

C.R.F.E

Access DB# 89382

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: M.A. WALICKA Examiner #: 78201 Date: March 19, 03
Art Unit: 1652 Phone Number 305-7270 Serial Number: 041833, 328
Mail Box and Bldg/Room Location: 10 D06 Results Format Preferred (circle): PAPER DISK E-MAIL
10 D01

If more than one search is submitted, please prioritize searches in order of need. MEJ

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Composition exhibiting a von Willebrand factor etc

Inventors (please provide full names): Saemle et al.

Earliest Priority Filing Date: 11/22/0000

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search SEQ ID NO: 1
SEQ ID NO: 15

and a fusion protein

SEQ ID NO: 1 + SEQ ID NO: 15
N → C

AA Thank you in advance,
1-12
15-136

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6806 TEL. (703) 305-1954

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: _____	NA Sequence (#) _____	STN _____
Searcher Phone #: _____	AA Sequence (#) <u>3</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>3/20</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>3/20</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: <u>10</u>	Fulltext _____	Sequence Systems _____
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: <u>10</u>	Other _____	Other (specify) _____

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:35:07 ; Search time 32 Seconds
(without alignments)
952.968 Million cell updates/sec

Title: SEQID_1_15FUSED
Perfect score: 773
Sequence: 1 AAGILHLELLVAVGPVQ.....SPTWSCLITDGFGLGVTI 148

Scoring table: BLOSUM62 Gapop 10.0 , Gapext 0.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacterioph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	773	100.0	1427	4 Q96L37	Q96L37 homo sapien
2	286	37.0	34350	4 Q8WZ42	Q8WZ42 homo sapien
3	272	35.2	26926	4 Q10466	Q10466 homo sapien
4	272	35.2	26926	4 Q8WZB3	Q8WZB3 homo sapien
5	264	34.2	9507	2 Q9EWA1	Q9EWA1 streptomyc
6	264	34.2	15281	3 Q09164	Q09164 tolypocladi
7	243	31.4	10917	2 Q93NW6	Q93NW6 streptomyc
8	241	31.2	9510	2 Q93NX9	Q93NX9 streptomyc
9	241	31.2	16215	5 Q9NFS3	Q9NFS3 drosophila
10	239	30.9	6889	16 Q8XS40	Q8XS40 raietonia s
11	238	30.8	10223	2 Q54296	Q54296 streptomyc
12	237	30.7	9376	2 Q95168	Q95168 pseudomonas
13	236	30.5	11096	2 Q9L4W3	Q9L4W3 streptomyc
14	233	30.1	7829	5 Q18559	Q18559 caenorhabdi
15	231	29.9	9477	2 Q9L4X3	Q9L4X3 streptomyc
16	231	29.9	17352	5 Q95YM2	Q95YM2 procambarus

17	229	29.6	2165	5 Q19791	Q19791 caenorhabdi
18	229	29.6	6396	2 Q9KID7	Q9KID7 streptomyc
19	229	29.6	7463	16 Q924X6	Q924X6 streptomyc
20	226	29.2	5532	2 Q9S0R4	Q9S0R4 streptomyc
21	225	29.1	6048	2 Q93H87	Q93H87 streptomyc
22	225	29.1	7160	5 Q23551	Q23551 caenorhabdi
23	224	29.0	4498	2 Q93HE5	Q93HE5 streptomyc
24	224	29.0	6685	12 Q91W06	Q91W06 transmissib
25	224	29.0	7107	5 Q9VAF7	Q9VAF7 drosophila
26	224	29.0	7576	2 Q9ZGA4	Q9ZGA4 streptomyc
27	223	28.8	1054	5 Q9W493	Q9W493 drosophila
28	222	28.7	6781	11 Q8VHN7	Q8VHN7 mus musculu
29	222	28.7	6781	12 Q91AV2	Q91AV2 porcine epi
30	221	28.6	6831	5 Q23550	Q23550 caenorhabdi
31	220	28.5	6146	2 Q93HJ5	Q93HJ5 streptomyc
32	220	28.5	6420	2 P95814	P95814 streptomyc
33	220	28.5	6875	6 Q28733	Q28733 cryptospori
34	220	28.5	8563	2 Q54297	Q54297 streptomyc
35	220	28.3	8817	2 Q53840	Q53840 polyangium
36	219	28.3	13055	5 Q09165	Q09165 caenorhabdi
37	218	28.2	5644	2 Q93MX8	Q93MX8 streptomyc
38	217	28.1	6307	4 Q8WKG9	Q8WKG9 homo sapien
39	217	28.1	8243	5 Q96554	Q96554 cryptospori
40	216	27.9	4845	11 Q88738	Q88738 mus musculu
41	216	27.9	5588	2 Q9ALM2	Q9ALM2 saccharopol
42	215	27.8	4349	4 Q9NYQ8	Q9NYQ8 homo sapien
43	215	27.8	4834	4 Q95714	Q95714 homo sapien
44	215	27.8	4898	2 Q93N87	Q93N87 streptomyc
45	214	27.7	4589	11 Q9WU10	Q9WU10 rattus norv

ALIGNMENTS

RESULT 1

Q96L37 PRELIMINARY; PRT; 1427 AA.

AC Q96L37;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Von Willebrand factor-cleaving protease precursor.
GN ADAMTS13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX PubMed=11557746;
RA Zheng X., Chung D., Takayama T.K., Majerus E.M., Sadler J.E., Fujikawa K.;
RT "Structure of von Willebrand factor-cleaving Protease (ADAMTS13), a Metalloprotease Involved in Thrombotic Thrombocytopenic Purpura.";
RL J. Biol. Chem. 276:41059-41063(2001).
DR EMBL; AY055376; AAL17652.1; -;
DR MEROPS; M12.241; -;
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn.MTpeptidse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tse_1; 4.
DR PROSITE; PS00315; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSPI; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Protease; Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 75 1427 VON WILLEBRAND FACTOR-CLEAVING PROTEASE.
SQ SEQUENCE 1427 AA; 153632 MW; EB1BC3AABCA4442 CRC64;

Query Match 100.0%; Score 773; DB 4; Length 1427;
Best Local Similarity 100.0%; Pred. No. 1.4e-29;

Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	AAGGILHLELLVAVGPDVQFQAQHQDTERYVLTNLNIGAEILLRDPSLGAQFRVHLVKMVL	60						
Db	75	AAGGILHLELLVAVGPDVQFQAQHQDTERYVLTNLNIGAEILLRDPSLGAQFRVHLVKMVL	134						
Qy	61	TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTDPGHADLVLYITRFDELPLDGNRQVRG	120						
Db	135	TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTDPGHADLVLYITRFDELPLDGNRQVRG	194						
Qy	121	VTOLGGACSPWTSCLITEDTGFDLGVTI	148						
Db	195	VTOLGGACSPWTSCLITEDTGFDLGVTI	222						
RESULT 2									
ID	Q8WZ42	PRELIMINARY;		PRT; 34350 AA.					
AC	Q8WZ42;								
DT	01-MAR-2002 (TReMBLrel. 20, Created)								
DT	01-MAR-2002 (TReMBLrel. 20, Last sequence update)								
DT	01-JUN-2002 (TReMBLrel. 21, Last annotation update)								
DE	Titin.								
CN	TTN.								
OS	Homo sapiens (Human).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
OX	NCBI_TaxID=9606;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=20309627; PubMed=10850961;								
RA	Freiburg A., Trombitas K., Hell W., Cazorla O., Fougereusse F.,								
RA	Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C.,								
RA	Granzier H., Labelit S.;								
RT	"Series of exon-skipping events in the elastic spring region of titin								
RT	as the structural basis for myofibrillar elastic diversity.";								
RL	Circ. Res. 86:1114-1121(2000).								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=21573839; PubMed=11717165;								
RA	Bang M.L., Centner T., Fornoff F., Geach A.J., Gotthardt M.,								
RA	McNabb M., Witt C.C., Labelit D., Gregorio C.C., Granzier H.,								
RA	Labelit S.;								
RT	"The complete gene sequence of titin, expression of an unusual -700								
RT	kda titin isoform and its interaction with obscurin identify a novel								
RT	Z-line to I-band linking system.";								
RL	Circ. Res. 89:1065-1072(2001).								
DR	EMBL; AJ277892; CAD12456.1; -.								
DR	InterPro; IPR000282; Cytok_receptor_2.								
DR	InterPro; IPR000719; Euk_pkinase.								
DR	InterPro; IPR000577; FGGY_kin.								
DR	InterPro; IPR003961; FN_III.								
DR	InterPro; IPR001092; HLH_basic.								
DR	InterPro; IPR003599; Ig.								
DR	InterPro; IPR003598; Ig_c2.								
DR	InterPro; IPR003006; Ig_MHC.								
DR	InterPro; IPR003596; Ig_v.								
DR	InterPro; IPR002016; Peroxidase.								
DR	InterPro; IPR004168; PPAK_motif.								
DR	InterPro; IPR002290; Ser_thr_pkinase.								
DR	Pfam; PF00041; fn3; 132.								
DR	Pfam; PF00047; Ig; 146.								
DR	Pfam; PF00069; pkinase; 1.								
DR	Pfam; PF02818; PPAK; 53.								
DR	ProDom; PD0000001; Euk_pkinase; 1.								
DR	SMART; SM00060; FN3; 133.								
DR	SMART; SM00409; IG; 167.								
DR	SMART; SM00408; IGC2; 148.								
DR	SMART; SM00406; IGV; 23.								
DR	SMART; SM00220; S_TKc; 1.								
DR	SMART; SM00219; TyrKc; 1.								
DR	PROSITE; PS00933; FGGY_KINASES_1; UNKNOWN_1.								

DR	PROSITE; PS00038; HELIX LOOP HELIX; UNKNOWN_1.								
DR	PROSITE; PS00290; IG_MHC; UNKNOWN_1.								
DR	PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.								
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; UNKNOWN_1.								
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.								
Qy	SEQUENCE 34350 AA; 3816262 MW; 5B1120058A7CE58A CRC64;								
Query Match 37.0%; Score 286; DB 4; Length 34350;									
Best Local Similarity 0.4%; Pred. No. 4.1;									
Matches 112; Conservative 11; Mismatches 24; Indels 29363; Gaps 32									
Qy	2	AGG-----	4						
Db	3131	AGGNVSTAKLFVEGRDVRIRSIKEVQVIEKQRAVVEFEVNEDDVDHAWYKDGIEINFQV	3190						
Qy	5	-----	4						
Db	3191	QERHKVVERIHRMFISETROSDAGEYTFVAGNRNRSSVTLVYNAPEPPQVLQELQPTV	3250						
Qy	5	-----	4						
Db	3251	OSGKPARFCAVIGRPOPQKISWYKEQLLSTGFKCKFLHDGQVYTLIIIEAPPEDAAVYT	3310						
Qy	5	-----	4						
Db	3311	CEAKNDYGWATTSASLSVPEVSPDQEMPVPPAIITPLQDVTTSQGQAPQCRVSG	3370						
Qy	5	-----	4						
Db	3371	TDLKVSWSKDKKIKPSRFRMTQFEDTYQLEIAEAYPEDEGTYTFVASNAGVQVSSSTAN	3430						
Qy	5	-----TLH-----LELL-----	11						
Db	3431	LSLEAPESILHERIDOEIEMEMKEPSSPSLAEELGSLHAELOLSKINETLELSESPVY	3490						
Qy	12	-----	11						
Db	3491	STKFDSEKGTGPIFIKEVSNADISMGDVATLSVTIVIGIPKIQWFFGVLLTPSADYK	3550						
Qy	12	-----	11						
Db	3551	FVFDGDDHSLIILFTKLEDEGEYTCNASNDYKTCISAYLKINSKGEGHKOTETESAVAK	3610						
Qy	12	-----	11						
Db	3611	SLEKLGPCPPHFLKELKPIRCAQGLPAIFEVTVGEPAPTVMFKNKQLCTSVYYTII	3670						
Qy	12	-----	11						
Db	3671	HNPNGSQTFIWDPQREDSGLYICKAENMLGESTCAEALLVLLEDMDTMTWPCAKSTPE	3730						
Qy	12	-----	11						
Db	3731	APEDFPOTPLKGAVALDSEQEIATFVKDTILKAALITEENQQLSYEHIKANELSSOL	3790						
Qy	12	-----	11						
Db	3791	PLGAQELQSILEQDKLTPESTREFLCINGSIHFOPLKEPSPNLQLOIVOSQKTFKSGIL	3850						
Qy	12	-----	11						
Db	3851	MPEPETQAVLSOTEKIFPMSAMSIEQINSITVEPLKTLAEPEGNYPOSSIEPPMHSYLT	3910						
Qy	12	-----	11						
Db	3911	SVABEVLSPKEKTVSDTNREQVTLQKQAQSALILSQSLAECHVESLOSPOVMISQVNY	3970						
Qy	12	-----	11						
Db	3971	EPLVPSEHSCTEGKILIESANPLENAGQDSAVRIBEGKSLRFPPLAEKQVLLKEHSD	4030						
Qy	12	-----	11						

DR	PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.								
DR	PROSITE; PS00290; IG_MHC; UNKNOWN_1.								
DR	PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.								
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.								
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.								
SQ	SEQUENCE 34350 AA; 3816262 MW; 5B1120058A7CE58A CRC64;								
Query Match 37.0%; Score 286; DB 4; Length 34350;									
Best Local Similarity 0.4%; Pred.No.4.1.1; Mismatches 24; Indels 29363; Gaps 32;									
Matches 112; Conservative 11;									
Qy	2	AGG-----	4						
Db	3131	AGGNTAKLFVEGRDVRIRSIKKEVQVIEKORAVVEFVNEDDVAHWYKDGIEINFQV	3190						
Qy	5	-----	4						
Db	3191	QERHKYVVERRIHRMFISETROSDAGEYTFVAGNRSSVLYVNAPEPOVLOEOPVTV	3250						
Qy	5	-----	4						
Db	3251	QSGKPARFCAVISGRPOPKISWYKEQLLSTGFKCKFLHDGOEYTLILLIEAFPDAAVTV	3310						
Qy	5	-----	4						
Db	3311	CEAKNDYGVAATTSASLSVEVPEVWSPQEMPVVPPAIITPLQDVTSEGPAPFOCRVSG	3370						
Qy	5	-----	4						
Db	3371	TDLKVSWSYKDKKIKRFRFRMTQFDTVQLIEAYPEDECTYTFVASNAVGVQSSTAN	3430						
Qy	5	-----ILH-----LELL-----	11						
Db	3431	LSLEAPESILHERIOEIEIEMEMKFPSSFLSAEEGLHSAELQLSKINETLELLSPVY	3490						
Qy	12	-----	11						
Db	3491	STKFDSEKGTGPIFIKEVSNADISMGDVATLSVTVIGIPKPIOWFNGVLLTPSADYK	3550						
Qy	12	-----	11						
Db	3551	FVFDGDHSLIILFTKLEDEGEYTCNASNDYKTCISAYLKINSKGEHKDTETESAVAK	3610						
Qy	12	-----	11						
Db	3611	SLEKLGPCPPHFKELKELPIRCAOGLPAIFVTVGEPAPTVTWKFNKQLCTSVYYTII	3670						
Qy	12	-----	11						
Db	3671	HNPNGSGTFIVNDPQREDSGLYICKAENMLGESTCAEELLVLEDTDMTTPCKAKSTPE	3730						
Qy	12	-----	11						
Db	3731	APEDFPOTPLKGPAAVEALDSEQEIATFVKDTILKAALITEENQQLSVEHIAKANELSSQL	3790						
Qy	12	-----	11						
Db	3791	PLGAQLQSLQDQLTPESTREFLCINGSIHFOPLKPSPNLQLOIVOSQKTFSEKIL	3850						
Qy	12	-----	11						
Db	3851	MPPEPTQAVLSDTEKIFPSPMSIEQINSLTVEPLKTLAEPSGNYPOSSIEPMSHYLT	3910						
Qy	12	-----	11						
Db	3911	SVAEVLSPKKTVDNTRQVTLQKQEAQSAIILSQSLAEGHVESLQSPDVMISQVNY	3970						
Qy	12	-----	11						
Db	3971	EPLVPSEHSCTEGKILIESANPLENAGQSAVRIBEGKSLRPLALEEKQVLLKEBHD	4030						
Qy	12	-----	11						

Db	4031	NVMPPOOIIESKREPVAIKKQVQGRDLLSKESLLSGIFEEQRLNKLQICRALQAAV	4090
Qy	12	-----	11
Db	4091	ASEQGLFSEWLNRIEKEVEAVNIQEPHRIMCMYLVTSAKSVTEVTIIIEDVDPQMA	4150
Qy	12	-----	11
Db	4151	NLKMELDALCAIYYEIDILTAEPRIQQAKTSLQEMDSFSGSKQVEPITEPEVSK	4210
Qy	12	-----	11
Db	4211	YLISPEVSYFNQSRVKYLDATPVTKGVASAVUSDEKQDESLKPEEKESSSESGETE	4270
Qy	12	-----	11
Db	4271	VATVKIQAEGGFIKEDGPMIHTPLVDTVSEBGDIVHLTTSITNAKEVNWYFENKLVPSD	4330
Qy	12	-----	11
Db	4331	EKFCLQDQNTYTLVIDKVNTEHOGYVCEALNDSGKTATSAKLTVVKRAAPVIKRKIE	4390
Qy	12	-----	15
Db	4391	PLEVALGHLAKFTCEIQSAPNVRQWFKAGREIYESDKCSIRSSKYISSLEILRTQVWDC	4450
Qy	16	-----	15
Db	4451	GEYCKASNEGVSCTATLTVTEAYPTFLSRPKSLTTFVGAAKFICTVTGTPTVIETI	4510
Qy	16	-----	15
Db	4511	WOKDGAALSPSPNWKISDAENKHILELSNLTIQDRGVYCKASKNKFADICQAEIILIDK	4570
Qy	16	-----	15
Db	4571	PHFIKELEPVQSAINKKHVLECOVDEDRKVTVTWSDGOKLPGDKYKICFEDKIATLEI	4630
Qy	16	-----	15
Db	4631	PLAKLDSGTGYCTAGNEAGSSCSATVTVREPPSFVKVDPSPYMLPGESARLHCKLKG	4690
Qy	16	-----	15
Db	4691	SPVIQVTFWKXNKELSESNTVRMYFVNSEAILDITDVKVEDSGSYCEAVNDVGSDCST	4750
Qy	16	-----	15
Db	4751	EIVIKEPPSFIKLEPADIVRGTNALLQCEVSGTGPEISWFKDKKQIRSSKKYRLFQK	4810
Qy	16	-----	15
Db	4811	SLVCLIEFSNADVGEYECVAVNEVGKCGCMATHLLKEPPTFVKVDDLIALGGQVTL	4870
Qy	16	-----	18
Db	4871	QAAVRGSEPISVTWKQGEVIREDKIKMSFNGVAVLIIPDVQISFGKYKTLAENEAG	4930
Qy	19	-----	18
Db	4931	SOTSVGELIVKEPAKIIERAELIOVTAGDPATLEVTVAGTPELKPKNYKDRPLVASKKY	4990
Qy	19	-----	18
Db	4991	RISFKNNVAQLKFYSAELHSDSQYTFEISNEVGSSCETTTFTVLDRDIAFFTYPRLNVD	5050
Qy	19	-----	18
Db	5051	SVVNOTCLDKIAGSLPMRVSWFKDGEIAASDRYIAFVEGTASLEIIRVDMNDAGNF	5110
Qy	19	-----	18
Db	5111	TCRATNSVSKDSSGALIQQEPPSPVTKPGSKDVLPGSAVCLKSTFGOSTPLTIRWFKGN	5170
Qy	19	-----	18
Db	5171	KELVSGSGCVITKEALESSELYLVKTSDSGTYTKVSNVAGGVECSANL FVKEPATFVE	5230
Qy	19	-----	18
Db	5231	KLPSQLLKKGDATQACKVTGTPPIKITWFANDREIKESSKURMSFVESTAVLRJLTDVG	5290
Qy	19	-----	18
Db	5291	IEDSGEYMCQAONEAGSDHCSSIVIVKESPYTKFEKPIEVLKEYDVMLLAEVAGTPPFE	5350
Qy	19	-----	18
Db	5351	ITWFKDNTILRSGRKYKTFIQDHLVSLQILKLVAAADAGEYOQCRVTNEVGSSICSARVTLR	5410
Qy	19	-----	21
Db	5411	EPPSFIKKIESTSLRGCTAAFOATLKGSLPITVTLKDSDEITEDDNI RMTFENNVA SL	5470
Qy	22	-----	21
Db	5471	YLSGIEVKHDGKYVCAKNDAGIQRCALLSVKEPATITBEAVSIDVTOGDPATLQVKFS	5530
Qy	22	-----	21
Db	5531	GTKEITAKWPKGOELTLGSKYKISVTDTVSILKIISTEKKDGEYTFEVQNDVGRSCK	5590
Qy	22	-----	21
Db	5591	ARINVLDLIIPSPFTKKLKKMDSIKGSFIDLECIVAGSHPISIQWFKDDQESASEKYKF	5650
Qy	22	-----	21
Db	5651	SFHDNTAFLEISQLEGDSGTYTCSATNKAGHNCOSGHLTVKBPYPYFVEKPSQDVNPT	5710
Qy	22	-----	21
Db	5711	RVOLKALVGGTAPMTIKWFKDKNELHGAARSVMKDDTSTLSLELFAAKATDSGTYICQLS	5770
Qy	22	-----	21
Db	5771	NDVGTATSKATL FVKEPPQFIKKPSVPLVLRNGOSTTFECQITGTPKIRVSWYLDGNEIT	5830
Qy	22	-----	21
Db	5831	AIQKHGISFIDGLATFQISGARVENSGTYVCEARNDA GTASCSIELKVKEPPTFIRELKP	5890
Qy	22	-----	21
Db	5891	VEWKYSDVELECEVTGTPPEVTWLKNNREIRSSKKYTLTDRVSVFNLHITKCDPSDTG	5950
Qy	22	-----	21
Db	5951	EYQIVSNEGSCSTRVALKEPPSFIKKIENTTTVLKSSATFQTVAGSPPISTMLK	6010
Qy	22	-----	21
Db	6011	DDQILDEDDNVVISFVDSVATLIQIRSDVNGHSGRYTQCAKNESGVERCYA FLLOEPAQI	6070
Qy	22	-----	21
Db	6071	VEKASVDVTEKDPMTLECVAGTPELKVWLKDGKQI VPSRYFMSFENNVA SFRIQSV	6130
Qy	22	-----	21
Db	6131	MKQDSQYTFKVENDFGSSCDAYLRLVLDQNI PPSFTKKLT KMDKVLGSSIHMECKVSGS	6190
Qy	22	-----	29
Db	6191	LPISAQWFKDGKEISTSAKYRLVCHERSVSLVNNLELEDTANYTCKVSNVAGDDACSGI	6250

QY 30 ----- 29
Db 6251 LTVKEPFLVKPGRQQAIPDSTVEPKAILKGTTPPKIKWFKDDDELVSQPKCFIGLEGS 6310
QY 30 ----- 29
Db 6311 TSFLNLYSDASKTGQYCHVTNDVSDSCTTLLVTEPPKFVKLEASKIVKAGDSSL 6370
QY 30 ----- 29
Db 6371 ECKIAGSPERVWFRNEHELPAISKYRMTFIDSVAVIQMNUNLSTEDSGDFICEAQNPAG 6430
QY 30 ----- 29
Db 6431 STSCSTKVIKPEPVFFSPPIYETLKNAEVSECELSGTTPPEVWVKDKRQLRSSKKY 6490
QY 30 ----- 29
Db 6491 KIASKNFHTSIHLNVDTSDIGYHCKAQNEVGSDTCVCTVKLKEPRFRFVSKLNSLTVA 6550
QY 30 ----- 29
Db 6551 GEPALQASIEGAQPIFVOWLKEKEVIRESENIRITFVENVATLQFAKAPANAGKYIC 6610
QY 30 ----- 29
Db 6611 QIKNDGGMEENWATLMVLEPAVIVEKAGPMTVTGETCTLECKVAGTPELSVEWYKDGKL 6670
QY 30 ----- 29
Db 6671 LTSSQKHSPYNKISSLRILSVRODAGTYTFOVQNNVGKSCTAVVDVSDRAVPPSFT 6730
QY 30 ----- 29
Db 6731 RRLKNTGGVLGASCILECKVAGSSPISVAPHEKTKIVSGAKYQTTFSDNVCTLQLNSLD 6790
QY 30 ----- 32
VLT-----
|||
Db 6791 SSDMGNYTCVAANVAGSDECAVLTVQEPSPFVKPEPELVLPQKNVTFTSVIRGTTPFK 6850
QY 33 ----- 32
Db 6851 VNWFRGARELVKGRCNIFYEDTVAELFNIDISQSEYTCVVSNAGQASCTTRLFVK 6910
QY 33 ----- 32
Db 6911 EPAFLKRLSDHSVEPGKSIILESTYTGTLPISTWKKDGFNITTSEKNIVTTEKTCIL 6970
QY 33 ----- 32
Db 6971 EILNSTKR DAGOYSCIEIENAGRDVCGALVSTLEPPYFVTELEPLEAAVGDVSLQCOVA 7030
QY 33 ----- 39
NLNIGAE-----
|||
Db 7031 GTPETVSWYKGDTKLRPTPEYRTFTNNVATLVFNKYNINDSGEYTCKAENSIGTASSK 7090
QY 40 ----- 43
LLRD-----
|||
Db 7091 TVFRIQEROLPPSFARQLKDIEQTGVLPTLTCLRLNGSAPIQVCWYRGDVLRRDENLOT 7150
QY 44 ----- 43
Db 7151 SPVDNVATUKILOTLDSHGQYSCSASNPLGTASSARLTAREPKSPFFDIKPVSDIVI 7210
QY 44 ----- 43
Db 7211 AGESADFECHVTGAQPMRITWSKDNKEIRPGGNYTITCVGNTPHLRLILKVGKDSGOYTC 7270
QY 44 ----- 43
Db 7271 QATNDVGKMCQAQSVKPEPFVKLEASKVAKOGESIQLECKISGSPKIVSWFRNDS 7330
QY 44 ----- 43

Db 7331 ELHESKWNYSFINSVALLTINEASABEDSGDYICEAHNGVDASCSTALTVKAPPVFQK 7390
QY 44 ----- 43
Db 7391 PSPVCAKAGSDVILQCEISGTPPEVWVVKDRKQVRNKKPKITSKHFDTSLHILNLEAS 7450
QY 44 ----- 43
Db 7451 DVGEYHCKATNEVGSDTSCSVKPEPRFVKLSLSDTSLIGDAVELRAIVEGQFQPSV 7510
QY 44 ----- 43
Db 7511 WLKDRGEVIRESENTRISFIDNIATLQLGSPASNSGKYICQIKNDAGMRECSAVLTVLE 7570
QY 44 ----- 43
Db 7571 PARIIEKPEPMTVTGTPFALECVVTGTPELSAKWKFKDRELSADSKHHITFINNKVASLK 7630
QY 44 ----- 43
Db 7631 IPCAEMSDKGLYSFEVKNSVGKSNCTVSVHVSNDRIVPSPFIRKLKDVNAILGASVVLSCR 7690
QY 44 ----- 43
Db 7691 VSGSAPISVGHFQDGNIEIVSGPKCQSSFSNVCTLNLSLLEPSDTGIYTCVAANVAGSDE 7750
QY 44 ----- 43
Db 7751 CSAVLTVQEPSPFQTPDSVEVLPGMSLTFTSVIRGTPPFVKWPKGSRRELVPGESCNIS 7810
QY 44 ----- 43
Db 7811 LEDFVTELELFEVQPLESGDYSLVNDAGSASCTTHLFVKPEPATFVKLADFSVETGSP 7870
QY 44 ----- 43
Db 7871 IVLEATYTGTPISVSWIKDEVLIQSBERCSITMTEKSTILESTIEDYVAQYSCLIEN 7930
QY 44 ----- 43
Db 7931 EAGQDICEALSVLEPPYFIEPLEHVEAVIGEPATLOCKVDGTPETIRISWYKEHTKLRS 7990
QY 44 ----- 43
Db 7991 PAYQMOQKNVASLVINKVDHSDVGEYSCKADNSVAVASSAVLVIKERKLPFFARKLK 8050
QY 44 ----- 43
Db 8051 DVHETLGPVPAPECRINGSEPLQVSWYKGVLLKDDANLQTSFVHNVAITMILQTDOSHI 8110
QY 44 ----- 43
Db 8111 GOYNCSASNPLGTASSAKLILSEHEVPPFPLKPVSVUDLALGESGTFKCHVTGTAPIKI 8170
QY 44 ----- 43
Db 8171 TWAKDNREIRPGGNYKMTLVNTATLTVLKVKGAGQYTCVASNAGKDCSAHLGVQE 8230
QY 44 ----- 43
Db 8231 PPRFIKLEPSRIVKQDEFTRECKIGSGSPEIKVLWYKDETEIQESSKFRMSPVDVAVL 8290
QY 44 ----- 43
Db 8291 EMINLSVEDSGDYTCNAHNAAGSASSTSLVKKEPPIFRKKPHPIETLKGADVHLECELO 8350
QY 44 ----- 43
Db 8351 GTPPFHVSWKDKRELRSKGYKIMSENFILTSIHLNVDAADIGEQCKATNDVGSDTCV 8410
QY 44 ----- 43

Db	8411	GSIALKAPPRFVKKLSDISTVVGKEVQLQTIEGAEPISVWVFXDKGEIVRESNMISY	8470
Qy	44	-----	43
Db	8471	SENIATLQPSRVEPANAGKYTCQIKNDAGMQECFATLSVLEPATIVEKPESIKVTTGTC	8530
Qy	44	-----	43
Db	8531	TLECTVAGTPELSTKWFKDGKELTSONKYKISFFNFKVSGLKLIINVPASDSGVYSEVQNP	8590
Qy	44	-----	43
Db	8591	VGKDSCTASLQVDRTPPSFTRKLTETNGLSGSSVWMECKVYGSPPISSVWFHEGNEIS	8650
Qy	44	-----	43
Db	8651	SGRKYQTLTDNTCALTVMNLEESDSDYTCIATNNAGSDECSAPLTVREPPSFVQKDDP	8710
Qy	44	-----	43
Db	8711	MDVLGTGNTVFTSIVKGTTPFSVSWFKSGSELVPGDRCNVSLSDSVAELELFDVDTSSQG	8770
Qy	44	-----	43
Db	8771	EYTCIVSNEAGKASCTHLYIKAPKFKVLNDYSIEKGKPLILEGTFGTTPPSIVTWKK	8830
Qy	44	-----	43
Db	8831	NGINVTPSQRCNITTTTEKSAILEPSSTVEDAQYNCYIENASGKDSQAQILILEPPYF	8890
Qy	44	-----	43
Db	8891	VKOLEPVKSVGDSASLQCLAGTPEIGSVWYKGTDKLRTTTYKMHFRNNVATLVFNQV	8950
Qy	44	-----	51
Db	8951	DINDSGEYICKAENSVEGSASTFLTVEQKLPSPFSRQLRDVQETVGLPVWFDCAISGS	9010
Qy	52	-----	51
Db	9011	EPISVSWYKDGKPLKDSPNVQTSFLDNTATLNFKTDRLAGQVSCATNPICGSASSAR	9070
Qy	52	-----	51
Db	9071	LILTEGNPPFFDIRLAPVDAVVGESADFECHVTGTQPIKVSNAKDSREIRSGGQYISY	9130
Qy	52	-----	51
Db	9131	LENSAHLTVLVDKDGDSQVTCYAVNEVGKDSCTAQLNIKERLIPPSTKRLSETVEETE	9190
Qy	52	-----	51
Db	9191	GNSFKLEGRVAGSQPITVANYKNNIEIOPTSNCEITFNKNTLVLOVRKAGNNDAGLYTCK	9250
Qy	52	-----	51
Db	9251	VSDNAGSALCTSSIVIKEPKPPVFDQHLTPVTVSEGEYVQLSCHVQSGSEPIRQWLKAG	9310
Qy	52	-----	51
Db	9311	REIKPSDRCSFASFAGTAVLELRDVAKADSGDYVCKASNVAGSDTTKSVTIKDKPAVAP	9370
Qy	52	-----	51
Db	9371	ATKKAADVGRLLFFVSEPOSIRVVEKTTATTAFAKVGCDPIPNVKWTKGKRWLNQNOGRVFI	9430
Qy	52	-----	51
Db	9431	HQKGDEAKLEIRDTTKTDSGLYRCVAFNEHGEIESNNLVQVDERKKQEKIEGDLRAMLKK	9490
Qy	52	-----	51
Db	9491	TPILKKGAGEBEIEDIMELLKNVDPKEYEKARMYGITDFRGLLOAFELLKQSQBEETHR	9550
Qy	52	-----	51
Db	9551	LEIEEIERSEDEKEPEELVSFIQORLSQTEPVTLIKDIENQTVLKDNDVAFEDIKINY	9610
Qy	52	-----	51
Db	9611	PEIKLSWYKOTEKLEPSDKFEISIDGDRHTLRVNCQLKDOGNRYRLVCGPHIASAKLTVI	9670
Qy	52	-----	51
Db	9671	EPAWERHLQDVTLKEGQCTCTMTQCFSPVNVKSEWFRNGRILKPOGRHKTEVEHKVHLTI	9730
Qy	52	-----	51
Db	9731	ADVRAEDQGYCTKYEDLETSABLRIEAEPIQTKRIQNIWSEHQSATPECEVSFDDAI	9790
Qy	52	-----	51
Db	9791	VTWYKGTTELTSOKYNFRNDGRCHYMTIHNVTDDGCVSVIARLEPRGEARSTABLYL	9850
Qy	52	-----	51
Db	9851	TTKEIKLELKPPDIPDSRVPIPTMPIRAVPPPEIIPVAVAPPILLPLTPPEKKPPPKRIE	9910
Qy	52	-----	51
Db	9911	VTKAVKDAKKVAKPKEMTPREEIVKKPPPTTLIPAKAPEIIDVSSKAAEEVKIMTIT	9970
Qy	52	-----	51
Db	9971	RKKEVQKEAEVYKQAVHKEKVFIESPEEPYDELEVEPYTEPPQYVEEDEDYEE	10030
Qy	52	-----	51
Db	10031	IKVEAKKEVHEEWEDEFEQGYEYEREBEGYDEGESEAEYQEREVIQVQKEVYESHER	10090
Qy	52	-----	51
Db	10091	KVPAKVPKAPPPPKVJIKKPVIEKTSRRMEEEKVQVTKVPEVSKKIVPQKPSRTPV	10150
Qy	52	-----	59
Db	10151	QEEVIEVKVPAVHTKEMVISEEKMFASHTEEVSVTVPEVQKEIVTEEKIHVAVSKRVE	10210
Qy	60	-----	59
Db	10211	PPPVPPELPEKPAPEEAVAPIPKKVPAPKVPVPPKVPPEKKVPVPKKEPAAPPK	10270
Qy	60	-----	59
Db	10271	VPEVPKVPPEEKIPVPVAKKEAPPAKVPVEQKRVVTEEKITIVTQREESPPPAVPEIP	10330
Qy	60	-----	59
Db	10331	KKVPEERKVPKKEEVPPPKVPALPKKVPPEKVAVPVPAKAKPPPAEVSCKTVV	10390
Qy	60	-----	59
Db	10391	EKKRFVAEEKSLFAVPQORVEVTRHEVSABEWSYSEEEBEGVSI SVYREEEEEEAEVT	10450
Qy	60	-----	59
Db	10451	EYEVMEEPBEYVVEEKHLIIISKRVEASPAEVTERQEKKIVLKPAPKIEEPPPAKVPEA	10510
Qy	60	-----	59
Db	10511	PKKIVPEKKVPAPVPKKEKVPKPPKVPPEKKVPKPKVPEKKVPKVKMBEPLPAKVTERHMQI	10570
Qy	60	-----	59
Db	10571	TOBEKVLVAVTKKEAPPKARVPEEPKRAVPEEKVLKLPKREEEPPPAKVTEFRKRVVKEE	10630

Db 12791 SREPKTFRHLKCTQBITGDDREPELIKDTGKHSWVIKSAAPFEDEAKYMFEDKHTSGKLI 12850
Qy 68 ----- 67
Db 12851 IEGIRLKFTPLKDVTAKEKESAVFTVELSHDNIRKWKFNQDRLHTTRSVSMODEGKTH 12910
Qy 68 ----- 67
Db 12911 SITFKDLSIDTSGIRVEAMGMSSEAKLTVLEGDPYFTGKLQDYTGVEKDEVLOCEISK 12970
Qy 68 ----- 67
Db 12971 ADAPVKWFKDGEIKPSKNAVIKADGKKRMLILKKALKSDIGQYTCDCGTDKTSGLKDIE 13030
Qy 68 ----- 67
Db 13031 DREIKLVRPLSHVEVMEETARFETEISEDDIHANWKLKGEALLOTPDCEIKEEGKIHSL 13090
Qy 68 ----- 67
Db 13091 VLHNCRLDOTGGVDFOANVKSANHLRVKPRVIGLLRPLKDVTVTAGETATFDCELSYED 13150
Qy 68 ----- 67
Db 13151 IPVEWYLGCKLEPSDKVVRSEGVHTLTLRDVKLEDAGEVQLTAKDFKTHANLFFVKEP 13210
Qy 68 ----- 67
Db 13211 PVEFTKPLEDQVEEGATAVLECEVSRENKVKWFKNGTEILKSKYIEVADGRVRKLV 13270
Qy 68 ----- 67
Db 13271 HDCTPEDIKTYTCDAKDFKTSNMLNVVPHVEFLRPLTDLQVREKEMARFECELSRENK 13330
Qy 68 ----- 67
Db 13331 VKWFKDGAIEIKKKYDIISKGAVRILVINKCLLDDEAEVSEVTRTSGMLTVLEBEA 13390
Qy 68 ----- 67
Db 13391 VFTKNLANIEVSETDTIKLVCEVSKFCAEVIWYKGBEIEIETGRYEILTEGRKRIILVION 13450
Qy 68 ----- 67
Db 13451 AHLEDAGNYNCRPLPSSRTDGKVKVHELAAEFISKPNLEILEGEKAEFFVCSISKESFPVQ 13510
Qy 68 ----- 67
Db 13511 WRDDKTLESGDKYDVIADGKRVLVVKDATLQDMGTVMVVGARAAAHLTVIEKLRIV 13570
Qy 68 ----- 67
Db 13571 VPLKTRVKEQQEVVFNCEVNTGAKAKWFRNEAIFDSSKYIILOKDLVTLIRDAHL 13630
Qy 68 ----- 67
Db 13631 DDQANYVSLNTHRGENVKSAANLIVBEEDLRIVEPLKDIETMEKKSVTFWCKVNRNLVT 13690
Qy 68 ----- 67
Db 13691 LKWTNGEEVPFONRVSRYVDKYKMLTIKDCGFPDEGEYIVTAGQDKSVAELLIIIEAPT 13750
Qy 68 ----- 67
Db 13751 EFVEHLEDQTVTEPDADVFCQLSREKANVWYRNGREIKEGKKYKFEKDGSIHRLIILKD 13810
Qy 68 ----- 67
Db 13811 CRLDDECEYACGVEDKSRARLFVEEIPVEIIRPPQDILEAPGADVFLAELNKKQVEVQ 13870
Qy 68 ----- 67
Db 13871 WLNNMNVVQGDHKHMMSEGIHRLQICDIKPRDQGEYRFTIAKDKEAPAKLELAAPKIK 13930

Qy 68 ----- 67
Db 13931 TADODLVVDYKGPLTMVVPYDAYPKAEAEWFKENEPLSTKTIOTTABQTSFRILEAKGD 13990
Qy 68 ----- 67
Db 13991 KGRYKIVLONKHGKAEFINLKVIDVPGVRNLEVTEFTFGEVSLAWEPLTDGGSKIIG 14050
Qy 68 ----- 67
Db 14051 YVVERRDIKRTWVLATDRAESCEFTVTGLQGGVEYLFVRSARNRVGTGEPVETDNVE 14110
Qy 68 ----- 70
Db 14111 ARSKYDVEGPPNLVTITDVRNRFVSLTWPEPEYDGGAEITNYVIELRDKTSIRMDTAMTV 14170
Qy 71 ----- ANLTSS-- 76
Db 14171 RAEDLSATVTDWEGOEYSFRVRAONRIGVCKPSAATPFVKVADPIERPSPVNLTSDDQ 14230
Qy 77 ----- 76
Db 14231 TOSSVQLKWEPLPKDGGSPILGYIIERCEEGKNWIRCNMKLVPELTVKVTGLEKNKYL 14290
Qy 77 ----- 76
Db 14291 YRVSANENKAGVSDPSEILGPLTADDAFVEPTMDLSAFKDGLEIVPNPITILVSPSTGYPR 14350
Qy 77 ----- 76
Db 14351 PTATWCFGDKVLETGDRVMKMTLSAYAEVLVSPSERSDKGIYITUKLENRVKTIISGEIDVN 14410
Qy 77 ----- 76
Db 14411 VIARPSAPKELKFGDITKDSVHLTWEPDDGGSPLTGYVVEKREVSRTWTKYMDFTVD 14470
Qy 77 ----- LLSVC----- 81
Db 14471 LEFTVDLVQGEYLFVKYCARNKGPGEPAYVDEPNVNMSTPATVDPDENVKMRDRTRANS 14530
Qy 82 ----- 81
Db 14531 IFLTWDPKNDGGRIKGYIIVERCPGRSDKKVACGEPVAETKMEVTGLEEGKWYAYRKA 14590
Qy 82 ----- 81
Db 14591 LNRQASKPSRPTIEIOAVDTQEAPEIFLDVKLLAGLTVKAGTKIELPATVTGKPEPKIT 14650
Qy 82 ----- 81
Db 14651 WTKADMILKQDKRITITENVPKSTVTIYDSKRSDTGTVIIIEAVNVCGRATAVVEVNVLDK 14710
Qy 82 ----- 81
Db 14711 PGPPAAFDITDVNESCCLLTWNPPRDDGSKITNYVERRATDSEVWHKLSSTVKDTNFK 14770
Qy 82 ----- 81
Db 14771 ATKLI PNKEYIIFRVAENMYGVGPVQASPI TAKYQFDPPGPPTRLEPSDITKDAVTLTW 14830
Qy 82 ----- 81
Db 14831 CEPDDDDGSPITGYWVERLDPTDKWRCNKMVPKDTTYRVKGLTNKKKYRFRVLAENLA 14890
Qy 82 ----- 81
Db 14891 GPCKPSKSTEPILIKDPTDPPWPGKPTVKOVGKTSVRLNMTKPEHDGGAIESYVIELM 14950
Qy 82 ----- 81
Db 14951 KTGTDENVRVAVGVPVTTQHLPLGLMEGOEYSFRVRAVNKAGESEPSDPVLCREKLYP 15010

Qy	82	-----	81
Db	15011	PSPRWLEVINITKNTADLKWTVPKXGGSPITNYIWEKEDVRKGMQTVDTTVKDTKCT	15070
Qy	82	-----	81
Db	15071	VTPLTEGSLVFRVAENAIGOSDYTEIEDSVLAKDTFTTPGPYALAVVDVTKRHVDLK	15130
Qy	82	-----	81
Db	15131	WEPPKNDGGRPIORYVIEKKERLGRVWKAGTAGPCNFRVTDVIEGTEVOFVRAENE	15190
Qy	82	-----	81
Db	15191	AGVCHPSETEILSIEDPTSPGPLDLHVTDAGRKHIAIAWKPEKNGGSPIIGYHVEM	15250
Qy	82	-----	81
Db	15251	CPVOTEKMRVNSRPIDLKFKVEGVDPKEYVLRVRAVNAIGVSEPSISENVVAKDP	15310
Qy	82	-----	81
Db	15311	DKPTIOLETHDIIVIEGEKLSIPVPFRVAVPVTVMHKGKEVKASDRLTMKNDHLSAH	15370
Qy	82	-----	81
Db	15371	LEVPKSVRADAGIYVITLENKLGSAASINVKVIGLPGCKDIKASDITKSSCKLTWEPP	15430
Qy	82	-----	81
Db	15431	EPDGGTPILVYLERREAGRRTYIPVMGENKLSWTVKDLIPNGEYFFRVKAVNKVGGGE	15490
Qy	82	-----	81
Db	15491	YIELKNPVIAQDPKQPPDPVDVEVHNPTAEAMTITWKPPLYDGGSKIMGYIEKIAKGE	15550
Qy	82	-----	81
Db	15551	ERWKRCHLVPIILTYTAKGLEGEKEYQFVRAENAAGISEPSRATPPTKAVDPIDAPKV	15610
Qy	82	-----	81
Db	15611	ILRTSVKRGDEIALDASISGSPYPTITWIKDENIVPVEIKKRAAPLVRRRKGEVQEE	15670
Qy	82	-----	81
Db	15671	EPFVLPLTQRLSDNSKKGESQLRVDRSLRDPDHGLYMIKVENDHGIAKAPCTVSVDTPG	15730
Qy	82	-----	81
Db	15731	PPINFVEDIRKTSVLCKWEPPLODDGSEIINYTLKKDKTKPDSEWIVTSTLRHCKYS	15790
Qy	82	-----	81
Db	15791	VTKLIEGKEYLFRVRAENRFGGPPCVSKPLVAKDPFGPDAPDKPIVEDVTSNMLVKW	15850
Qy	82	-----	81
Db	15851	NEPKONGSPILGWLEKREVNSTHWSRVNKSLLNALKANVDGLLEGLTYVFRVCAENAAG	15910
Qy	82	-----	81
Db	15911	PGKSPSPDPKTAHDPIGPPGPIPRVTDTSSTTIELEWEPPAFNGGGEIVGYFVDKOLV	15970
Qy	82	-----	81
Db	15971	GTNEWSRCTEKMIKVROYTVKREIAREGADYKLRVSAVNAAGEPPGPGTOPVVAEPQEP	16030
Qy	82	-----	81
Db	16031	VELDVSVKGGIOIMAGKTLRIPAVVTGRPVPTKWTKBEGELDKORVVIDNVGTSKELII	16090
Qy	82	-----	81

Db	16091	KDALRKOHGRYVITATNSCGSKFAAARVEVFDVPGPVLDLKPVVVTRKMKCLLNWSDPEDD	16150
Qy	82	-----	81
Db	16151	GGSEITGFIIEKDKAMHTWRQPIETERSKCDITGLLEGOBYKPRVIAKNKFGCGPPVEI	16210
Qy	82	-----	81
Db	16211	GPILAVDPLGPTSPERLTYTERTKSTITLDWKEPRNGSGSPIQGYIIEKRRHDKDPFER	16270
Qy	82	-----	81
Db	16271	VNKLCPPTSFLENLDEHOMYEPRVKAINEIGESEPSLPLNVVIQDDEVPTIKRLSV	16330
Qy	82	-----	81
Db	16331	RGDTIKVKAGEPVHVPADVTLGLPMPKIEWSKNETVIEKPTDALQITKEEVSRSKTELS	16390
Qy	82	-----	81
Db	16391	IPKAVREDKGTVTYTASNRLGSVFRNVHVEVYDRPSPPRNLAVTDIKAESCYLTDAPLD	16450
Qy	82	-----	81
Db	16451	NGGSEITHYVIDKRDA SRKAEWEVNTAVEKRYGIWKLPNGQYEPRAVNKYGISD	16510
Qy	82	-----	81
Db	16511	ECKSKWVIQDPYRLPGPPGPKVLARTKGSMLVSWTPLLNGGSPITGYWLEKREBGP	16570
Qy	82	-----	81
Db	16571	YMSRVSRAPITKVGLUGVEFNVPRLLLEGVKYQFRAMAINAAGIGPPSPSEVAGDPF	16630
Qy	82	-----GW-----	83
Db	16631	PPGPPSCPEVKDKTKSSISLGMKPPAKDGGSPIKGYIVEMOEBEGTTDKRVNPEDKLITT	16690
Qy	84	-----	83
Db	16691	CECVVNLKELRYKRVKAVNAEGESEPSDTTGEIPATDIOEEPEVFIDIGAQCCLVCK	16750
Qy	84	-----	83
Db	16751	AGSQIRIPAVIKGRPTPKSSWEFDGKAKKMDGVHDIPEDAQLETAENSSVIIIECKR	16810
Qy	84	-----	83
Db	16811	SHTGKYSITAKNAGQKANTANCRVKVMDVPGPKDLKVSIDITRGCRLSHMPDDGGDRI	16870
Qy	84	-----	83
Db	16871	KGYVIEKRTIDGKAMTKVNPDCGSTTFFVVDLLSEQQYFFRVRAENRFOIGPPVETIQRT	16930
Qy	84	-----	83
Db	16931	TARDPIYPPDPPIKLKIGLITKNTVHLSWKPPKNDGSPVTHYIVECLAWDPTGTTKKEAW	16990
Qy	84	-----	83
Db	16991	ROCNRDVEELOFTVEDLVEGGEYEPVKA VNAAGVSKPSATVGPCCORPDMPPSIDLK	17050
Qy	84	-----	83
Db	17051	EFMEVEEGTNVNIIVAKIKGVPPFTLTWFKAPPKKPNKEPVLVDTHVKNLVDDTCTLVI	17110
Qy	84	-----	83
Db	17111	PQSRSDTGLYTTITAVNNLGTASKEMRLNVLGRPGPPGVPIKFESVSAQMTLSWFPKPD	17170
Qy	84	-----	83

seqid_1_15fused.rspt

Thu Mar 20 13:50:31 2003

Db	17171	DGSKITNYVIEKREANRKTWHVSSPEKCTYTIPIKLEGEHYFRIMAQNKYGIGEP	17230
Qy	84	-----SOTIN-----	88
Db	17231	DSEPETARNLFSVPCAPDKPTVSSVTRNSMTVNWEEPEYDGGSPVTGYWLEMKDITTSKR	17290
Qy	89	-----	88
Db	17291	KRVNRDPIKAMTLGVSQYKVTGLIEGSDYQFRVYAINAAGVGPASLPSPATARDPIAPP	17350
Qy	89	-----	88
Db	17351	PFPKVTMDTKSSADLEWSPLPKDGGSKVTGYIYEKSEGEWEKDKVGRGKLVVT	17410
Qy	89	-----	88
Db	17411	GLKEGAFYKFRYSAVNIAGIGEPGEVDVEMKDRLVSPDLQDASVRDRIVVHAGVIR	17470
Qy	89	-----	88
Db	17471	IYVSGKPPPTVTNMNERTLPOEATIIETTAISSMWIKNCORSHQGVYSLAKNEAGE	17530
Qy	89	-----PEDD-----	92
Db	17531	RKKTIIIDVLDPGPGVGTFFLAHNLNESCULTWFSPEDDGGSPITNYVIEKRESDRRAW	17590
Qy	93	-----	92
Db	17591	TPVTYVTRQNATVQGLIOCKAYFRIAAENSGMGPVETSEALVIREPITVPERPED	17650
Qy	93	-----	92
Db	17651	EVKEVTNTVTLWNPYPKVDGSGEIIINYLESRLIGTEKFKVTNDNLLSRKYTVKGLKE	17710
Qy	93	-----	92
Db	17711	GDTYEVRSVAVNIYGOKPSFCTKPTCKDELAPPTLHLDKLTIRVGEAFALTGRYS	17770
Qy	93	-----	92
Db	17771	GKPKVSWFKBADVLEDDRTHIKTTPATLALEKIKAKRSDSGKCVVVENSTGSRKGF	17830
Qy	93	-----	92
Db	17831	QVNVDRPGPVPVGSFDEVTQDMVISMKPPDDGGSKITNYIIEKKEVGKDVMPVT	17890
Qy	93	-----	92
Db	17891	SASAKTTCKVSKLEKGDYIFRIHAENLYGISDPLVSDSMKAKDRFRVPDAPDQPIVEV	17950
Qy	93	-----	92
Db	17951	TKDSALVTWKNPHDGGKPIITNYILEKRETMSKRWARTKOPHPTKFRVPDLLEGQYE	18010
Qy	93	-----	92
Db	18011	PRVSAENEIGDPSPSKPVFAKDIAPKPSPPVPEAIDTTCNSVLTWQPPRHGGSK	18070
Qy	93	-----TDPGH-----	97
Db	18071	ILGYIVEYQVGBEWRANHTPESCPTYKYVTGLRDGQYKPRVLAVNAAGESDPAHV	18130
Qy	98	-----	97
Db	18131	PEPVLVKDRLEPPELILDANWAEQHIKVGDTLRLSAILKGVPPKVTWKEDRDAPTKA	18190
Qy	98	-----	97
Db	18191	RIDVTPVGSKLEIRNAHEDGGIYSLTVENPAGSKTVSKVLVDKPGPPRDLEVSEIRK	18250
Qy	98	-----	97
Db	18251	DSCYLTKWEPLDDGGSVITNYVVERRDVASQWSPLSATSKKSHFAKHLNEGNOYLFV	18310
Qy	98	-----	97
Db	18311	AAENQYGRGPFVETPKPIKALDPLHPPGPKOLHHDVVDKTEVSLVMNKPDRDGGSPITG	18370
Qy	98	-----	97
Db	18371	YLVEYOBEGTQDWIKFKFTVTNLECVVTGLOOGKTYFRVKAENIVGLGLPDTTPIECOE	18430
Qy	98	-----	97
Db	18431	KLPPPSVELDVKLLIEGLVVKAGTTVRFPATIRGVVPVPTAKMTTGDSEIKTDEHYTVETDN	18490
Qy	98	-----	97
Db	18491	FSSVLTIKNCRLRDTGEYQITVSNAGSKTVAVHLTVLDVPGPPTGPINILDVTPHEMTI	18550
Qy	98	-----	97
Db	18551	SWOPPKDDGGSPVINYIVEKODTRKDTGWGVVSSGSSKTKLKI PHLOKGEVVFVRAENK	18610
Qy	98	-----	97
Db	18611	IGVCPPLDSTPTVAKHKFSPSPGPKPVVTDITENAAATVSWTLPSDGGSPITGYYMERR	18670
Qy	98	-----	97
Db	18671	EVTGKWRVVKTPIDLDKFRVTGLYEGTVEFRVFAENLAGLSPSSSPDIKACRPIKP	18730
Qy	98	-----ADLV-----	101
Db	18731	PGPPINPKLKDOKSRETADLVWTKPLSDGGSPILGYVVECKPGTAOMNRINKDELIROCA	18790
Qy	102	-----	101
Db	18791	FRVPLGIEGNEYRFRKAANIVGEPRELAESVIAKDILHPPEVELDVTCRDVITVRVG	18850
Qy	102	-----	101
Db	18851	QTIIRILARVKRPEPDIITWTKGKVLVREKRVLDIQLPRVELOIKEAVRADHKYIISA	18910
Qy	102	-----	101
Db	18911	KNSSHAOGSAIVNVLDPRPGQCNLKVNTVKENTISWENPLDNGSGSEITNFIVEYRKP	18970
Qy	102	-----	101
Db	18971	NQKGSIIVASDVTKRLIKANLLANNEYIFRVCAENKVGVGPTIETKTPILAINPIDRPG	19030
Qy	102	-----	101
Db	19031	PENLHIADGKTFVYLKWRRPDYDGGSPNLSYHVPERLKGSDDERVHKSGIKETHYMYD	19090
Qy	102	-----	101
Db	19091	RCVENQIYEFVQTKNEGESDMVKTEVVVKEDLQKPVLDLKLSGVLTAKAGDTIRLEA	19150
Qy	102	-----	101
Db	19151	GVRGKPFPEVAWTKOKDATDLTRSPRVKIDTRADSKFSLTKAKRSDGKGYVVTATNTAG	19210
Qy	102	-----	101
Db	19211	SFVAYATVNVLDKPGPVRLKIVDVSSDRCTVCMWDPEDDGGCEIQNYILEKCEKRMVW	19270
Qy	102	-----	101
Db	19271	STYSATVLTPTGTTVTRLIEGNEYIFRVRAENKIGTPPTESKPVIAKTKYDKPRDPPE	19330
Qy	102	-----	101
Db	19331	VTKVSKEMTVVMNPPEYDGGKSIITGYFLEKKEKHSTRVFNKSAIPERRMKVQNLDP	19390

Qy	102	-----	101
Db	19391	HEYQFRVKAENEIGIGEPSLPSRPVAKDPIEPGPPNFRVVDTTKHSITLWGKPVYD	19450
Qy	102	-----	101
Db	19451	GGAPIIGYVVMRPKIADASPDEGKRCNAQAOLVRKEFTVTSLENQEYEFVCAQNOV	19510
Qy	102	-----	101
Db	19511	GICRPALKEAIKPKLEIPPEIDLDASMRKLVIVRAGCPILRFIAIVRGPRAPKVTMRKV	19570
Qy	102	-----	101
Db	19571	GIDNVRKGQVLDVTMAFLVIPNSTRDSDKYSYLTIVNPAGEKAVFVNVRLDTPGPVS	19630
Qy	102	-----	101
Db	19631	DLKVSVDTKTSCHVSWAPPENDGGSQVTHYIVKEADRKWTSTVTPVKKTSFHVTNLV	19690
Qy	102	-----	101
Db	19691	PGNEYFRVTAVNEYGPGVPTDVPKVLASDPLSEPPDPKLEVTMTKNSATLAWLPL	19750
Qy	102	-----	101
Db	19751	RDGAKIDGYITSYREBEQPADRWTEYSVVKDLSLVTVGLKEGKKYKFRVAARNAGVSL	19810
Qy	102	-----	101
Db	19811	PREAGVYEAKELPPKILMPEQITIKAGKKLRIEAAHVYKPHPTCKWKKEDEVVTSS	19870
Qy	102	-----	101
Db	19871	HLAVHKADSSILIIKDVTBKDSGYSLTAENSGTDTQIKVYVMDAPGPPPPFDISD	19930
Qy	102	-----	101
Db	19931	IDADACSLSWHPILEDGSGNTNIVIEKCDVSGDWNTALASVTKTSRVGKLIPOGEYI	19990
Qy	102	-----	101
Db	19991	FRVRAENRFGISEPLTSPKMAQPPFGVPSEPKNARVTKNKDCIFVAMDRPDSGSGPI	20050
Qy	102	-----	101
Db	20051	IGYLIKERNSLLWVKANDTLVRSTEYPCAGLVEGLEYSFRIYALNKAAGSSPPSKPTEY	20110
Qy	102	-----	101
Db	20111	VTARMPVDPGKPEVIDVTSTVSLIWARPKHDGSGKIIGYFVEACKLPGDKWRCNTAP	20170
Qy	102	-----	101
Db	20171	HQIPOEYATGLEEKAQYQFPAIARTAVNISPPSESDPVTILAENVPPIRIDLSVAMKS	20230
Qy	102	-----	101
Db	20231	LLTVKAGTNVCLDATVFGKPMPTVSKWKGDTLLKPAEGIKNAMQORNLTLELFSVNRKDS	20290
Qy	102	-----	101
Db	20291	GDYTTAENSSGKSATIKLVLDKPGPPASVKINKMYSDRAWLSWEPLEDGSGSEITNY	20350
Qy	102	-----	101
Db	20351	IVDKRETSRPNQAQVSATVPITSCSVEKLIEGHEYOFRICAENKYGVSDPVFTEPAIAKN	20410
Qy	102	-----	101
Db	20411	PYDPPGRCDDPPVISNITKDHMTVSWKPPADGSGPITGYLLEKRETOAQVNTKYNRPDI	20470
Qy	102	-----	101

Db	20471	ERTLKATGLQEGTEYEFVRVTAINKAGPGKPSDASKAAYARDPOYPAPPAPPKVYDTRTS	20530
Qy	102	-----	101
Db	20531	SVLSLWGKPAYDGGSPIIGYLVKVRADSDNWRNCLPQNLOKTRFEVTGLMEDTOYQFR	20590
Qy	102	-----	101
Db	20591	VYAVNKIGYSDPSDVPDKHYPKDILIPPEGELDADLRLKTLILRAGVTMRLVVPVKGRPP	20650
Qy	102	-----	101
Db	20651	KITWSPNVNLRDRIGLDIKSTDFDTFLRCENNVNKYDAGKYILTLENSCGKKEYTIVVKV	20710
Qy	102	-----	101
Db	20711	LDTGPPVNVTVKEISKDSAYVTWEPPIIDGSPPIINVVQKRDARSKWSVTTECSKT	20770
Qy	102	-----	101
Db	20771	SFRVANLEBEGKSYFPRVFAENEYGIGDPGETRDAVKASOTPGPVVDLKVRSVSKSSCSIG	20830
Qy	102	-----	101
Db	20831	WKKPHSDGSRIGYVVDFLTEENKQRMKSLSLQYSAKOLTEGKEYTFRVSAENGE	20890
Qy	102	-----	101
Db	20891	GTPSEITVVARDDVVAPODLKGLDPLCYLAKENSFRLKPIKPKPAPSVWKKGEDPL	20950
Qy	102	-----	101
Db	20951	ATDTRVSESSAVNTLIVYDCQKSDAGKYTITLKNVAGTKEGTISIKNVKGPIPTGPI	21010
Qy	102	-----	101
Db	21011	KFDEVTAEAMTLKWAAPPKDDGSEITNVILEKRDVNNKWNVCASAVOKTTFPRVRLHEG	21070
Qy	102	-----	101
Db	21071	MEYTRVSAENKYGVGEGLKSEPIVARHPFDVPDAPPPNIVDVRHDSVSLTWTDPKKTG	21130
Qy	102	-----	101
Db	21131	GSPITGYHLEFKERNSLWKRANKTPIRMDFKVTGLTEGLEYSFVPMALNAGVKPSEL	21190
Qy	102	-----	101
Db	21191	PSEPVALDPIDPPGKPEVINITRNSVTLIWTEPKYDGGHKLGTGYIVEKGLPSKSWMKA	21250
Qy	102	-----	101
Db	21251	NHNVPECAFTVTDLVEGGKYEFRIRAKNTAGAISAPSESTETIICKDEYEAPTIVLDP	21310
Qy	102	-----	101
Db	21311	IKDGLTIKAGDTIVLNAISILGKPLPKSSWAKGDIRPSDITQITSTSTSSMLTIKYAT	21370
Qy	102	-----	101
Db	21371	RKDAGEYTTATNPPFGTKVEHVKVTVLDVPGPPGVEISNVSAEKATLTWTTPLEDGGSP	21430
Qy	102	-----	101
Db	21431	IKSVILEKRETSRLMTVWSEDIQSCRHVATKLIQNEVIFRVSAVNHYGKGPVSEPV	21490
Qy	102	-----	101
Db	21491	KWVDRFGPPGPEKPEVSNVTKNATVSWKRPVDDGSGEITGVHVERREKSLRWRAIK	21550
Qy	102	-----	101

Db	21551	TPVSLRCKVTGLQEGSTYEFVRSAENRAGIGPSEASDSVLMKDAAYPPGPPSPHVTD	21610
Qy	102	-----LYTRF-----	107
Db	21611	TTKKSASLAWKPHYDGGLEITGYVVEHQVGEAWIKDTTGTALRIITQFVVPDLQKEX	21670
Qy	108	-----	107
Db	21671	YNFRISAINDAGVGEPAVIPDVEIVEREMAPDFELDLRLTLVVRAGLSIRIFVPIKGR	21730
Qy	108	-----	107
Db	21731	PAPEVTWTKDNINLKNRANIENTESFTLLIIPECNRYDTGKFVMTIENPAGKKSGFVNR	21790
Qy	108	-----	107
Db	21791	VLDTPGVLNLRPTDITKDSVTLHMDPLIDGSRITNIVIEKREATRKSSTATTCKHK	21850
Qy	108	-----	107
Db	21851	CTYKVTGLSEGCEYFPRVMAENEYIGIGETPETTEPVKASEAPSPDSLNMIDITKSTVSL	21910
Qy	108	-----	107
Db	21911	ANPKPKHGGSKITGYVIEAQRKSGDQWTHITTVKGLECVNRNLTEGEEYTFQVMVNSA	21970
Qy	108	-----	107
Db	21971	GRSAPRESRPIVKEQTMPELDLGIYOKLVIKAGDNIKVEIPVLGRPKPTVTWTKGD	22030
Qy	108	-----	107
Db	22031	QILKQORVNFETATSTILNINECVRSDSGPYPLTARNIVGEVDVITIQVHDIPGPT	22090
Qy	108	-----	107
Db	22091	GPIKFEVSSDFVTFSDPNDGGVPISNYVVMRQDSTTWVELATTVIRTYKATRL	22150
Qy	108	-----	107
Db	22151	TTGLEQYQVRKAQNRVGVGPGITSACIVANYPFKVPGPPTQVAVTKDSMTISWHEPL	22210
Qy	108	-----	107
Db	22211	SDGSPILGYHVERKERNGILWQVSKALVPGNIFKSSGLTDGIAYEFVRVIAENMAGSK	22270
Qy	108	-----	107
Db	22271	PSKPSERMLADPIDPPCKPVLNITRHTVTLKWAKEPYTGGFKITSYIIVEKRDLPNGRW	22330
Qy	108	-----	107
Db	22331	LKANFSNILENFTVSLGTEDAAYEFVRVIAKNAAGAISSPSEPSDAITCRDDVEAPKIV	22390
Qy	108	-----	107
Db	22391	DVKFKDVTILKAGEAFRLVADSVGRPPPTMEWSKDGELEGTAKLEIKIADFSTNLVND	22450
Qy	108	-----	107
Db	22451	STRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPEGLAVTEVTSEKCVLSWPPPLDDG	22510
Qy	108	-----	107
Db	22511	GAKIDHIYVQKRETSRLAWNVASEVQVTKLVTKLKNGEYIFRVMAVKNYGVGEPLS	22570
Qy	108	-----	107
Db	22571	EPVLAVNPYPPDPKPNPEVTTITKDSMVVCMGHDPDSGSGEIIINYIIVERRDKAGORWIK	22630
Qy	108	-----	107
Db	22631	CNKRTLILRKYVSGLTEGHEGYEFIRMAENAGISAPSTSPFYKACDVTFKPGPPGNR	22690
Qy	108	-----	107
Qy	108	-----	107
Db	22691	VLDTRSSISIAMNKPIYDGSBITGYMVEIALPEEDHQIVTPPAGLKATSYTITGLTE	22750
Qy	108	-----	107
Db	22751	NOBYKIRIYAMNSEGELGEPALVPGTPKAEDRMPLPPEILDADLRKVVITIRACCTLRLFVP	22810
Qy	108	-----	107
Db	22811	IKGRPAPEVKWARDHGESLDKASIESSTSYLLIIVGNVNRFDGKVIITVENSNGSKSAF	22870
Qy	108	-----	107
Db	22871	VNVRVLDTPGPPDLKVEVTKTSVTLTMDPPLLDGSGSKINIVIEKRESTRKAYSTVAT	22930
Qy	108	-----	107
Db	22931	NCHKTSWKVDOLQEGCSYFRVLAENEYIGLPAETAESVKASERPLPPGKITLMDVTRN	22990
Qy	108	-----	107
Db	22991	SVSLSWEKPEHGGSRILGYIVEMQTKGSKMATCATVKVTEATITGLIOGEEYSFRVSA	23050
Qy	108	-----	107
Db	23051	QNEKGISDPROLSVPVIAKDLVIPAPKLLFNTFTVLAGEDLKVDPFIFGRPTPAVTHHK	23110
Qy	108	-----	107
Db	23111	DNVPLKQTRVNAESTENNSLLTIKDACREDVGHVVVKLTNSAGEAETLNVILVDKPGP	23170
Qy	108	-----	107
Db	23171	PTGPVMDVETADSTILSWGPPKYDGGSSINNYIIVEKEDTSTTTWQIVSATVARTTIKAC	23230
Qy	108	-----	107
Db	23231	RLKTCEYQFRIAENRYGKSTYLNSEPTVAQYFPKVPGPPTPVVTLSSRDSMEVOWNE	23290
Qy	108	-----	107
Db	23291	PISDGSRVIGYHLERKERNSTLWVKLNKTPIPQTKFTTGLEGEVGEYEFVRSAENIVGI	23350
Qy	108	-----	107
Db	23351	GKPSKVSECYVARPCDPPGRPEAIIVTRNSVTLQWKXPTYDGGSKIYGYIIVEKKELPEG	23410
Qy	108	-----	107
Db	23411	RMKASFTNIIIDTHFEVTCGLVEDHRYEFVRVIAARNAAGVPSESESTGAIARDEVDPRI	23470
Qy	108	-----	107
Db	23471	SMDPKYKDTIVVHAGESFKVDADYKGPITQIWIQKGOELSNARLEIKSTDFATSLV	23530
Qy	108	-----	107
Db	23531	KDAVRVDSGNYILKAKNVAGERSVTNVKVLDRPGPPEGPVVISGVTAEKCTLAWKPPLQ	23590
Qy	108	-----	107
Db	23591	DGSDIINYIIVERRETSRLVWTVVDANVQTLCKVTUKLLEGNEYTFRIMAVKNYGVGEPL	23650
Qy	108	-----	107
Db	23651	ESEPVVAKNPFVVPDAPKAVEVTTVTKDSMIVVWERPASDGGSEIILGYVLEKRDKEGIRW	23710
Qy	108	-----	107
Db	23711	TRCHKRLIGELRLRVTLGLENHDYEFVRSAENAGLSEPPSAYQKACDPIYKPGPPNN	23770

Qy 108 ----- 107
Db 23771 PKVIDITRSSVFLWSKPIYDGGCEIQYIVERKCDVSVGEWMTCTPTTGINKTNI EVEKL 23830
Qy 108 ----- 111
Db 23831 LEXHEYNFRICAINKAGVGEHADVPGPPIVEEKLEAPDIDLDLELRKIINIRAGGSLRUF 23890
Qy 112 ----- 111
Db 23891 VPIKGRPTPEVKMGKVDGEIRDAAIIDVTSSFTSLVDNVNRYDSGKYTLTLENSSGTKS 23950
Qy 112 ----- 111
Db 23951 AFVTVRLDTPSPVNLKVTETIKDSVSITWEPPELLDGGSKIKNYIVERKEATRKSYAAV 24010
Qy 112 ----- 111
Db 24011 VTNCHKNSWKIDLOEGCSYFRVTAENEYIGLPAQTADPIKVAEVPQPPGKITVDVDT 24070
Qy 112 ----- 111
Db 24071 RNSVSLSWTKPEHGGSKIIQYIVEMOAKHSEKWCARVKSLOAVITNLTOGEEVLFV 24130
Qy 112 ----- 111
Db 24131 VAVNEKGRSDPRSLAVPIVAKDLVIEPDVKPAFSSYSVQVQODLUKIEVPISGRPKPTITW 24190
Qy 112 ----- 111
Db 24191 TKDGLPLKQTRINVTDSDLTLTILSIKETHKDDGGQYITVANVVGOKTASIEIVTLDKP 24250
Qy 112 ----- 111
Db 24251 DPPKGVKFDVSAESITLSNPLTYGGCOITNVIQKRDTTTTVMDVWSATVARTLK 24310
Qy 112 ----- 111
Db 24311 VTKLKTGYQFRIFAENRYGQSALESDDPIVAQPYKEPGPGPTPATAISKDSMWIQW 24370
Qy 112 ----- 111
Db 24371 HEPVNGGSPVIGYHLERKERNILWTKVNTIIHDTQFKAQNEEGIEYEFVVAENIV 24430
Qy 112 ----- 111
Db 24431 GVGKASKNSECYVARDPCDPPGTPPEIMVKRNEITLOWTKPVYDGGSMITGYIVEKRDLP 24490
Qy 112 ----- 111
Db 24491 DGRMMKASFTNVIQFTVSLGTEDQRYEFVRVIAKNAAGAIKSPSDSTGPITAKDELPL 24550
Qy 112 ----- 111
Db 24551 RISMDPKFRDITVNAVGETFRLEADVHGKPLPTIEWLRGDKEIESARCEIKNTDFKALL 24610
Qy 112 ----- 122
Db 24611 IVKDAIRIDGGQYILRASNAVAGSKFPVNVKVLDRPPEGPVQVGTVTSKCSLTWSPP 24670
Qy 123 -OLGG- ----- 126
Db 24671 LDGGSDISHYVVEKRETSRLANTVVADEVVNTSLKVLLEGVEYVFRIMAVNKYGVGE 24730
Qy 127 ----- 126
Db 24731 PLESAPVLMKNPVLPGPPKSLVNTIAKDSMTVCNWRPDSGGSEIIIGYIVEKDRSGI 24790
Qy 127 ----- 130
Db 24791 RWIKCNKRITDLRLVTLGTEDHEYEFVRSNAAGVGPSPATVYKACDPVFKGPP 24850
Qy 131 ----- 130

Db 24851 TNAHIVDTTKNSITLAWGKPIYDGGSEILGVVWEICKADEEEOIIVTPQOTGLRTRFEIS 24910
Qy 131 ----- 130
Db 24911 KLTEHQEYKIRVCALNKVGLGEATSVPGTVKPEDKLEAPELDLDSELKGIIVRAGGSAR 24970
Qy 131 -----TWS----- 133
Db 24971 IHIPKGRPTPEITHSREGEFTDKVQIEKGVNYTQLSIDNCDRNDACKYILKLENSSGS 25030
Qy 134 ----- 133
Db 25031 KSAFVTVKVLDTGPPQNLAVKEVRKDSAFVWEPPIIDGGAKVKVYVIDKRESTRKAYA 25090
Qy 134 ----- 133
Db 25091 NVSSKCSKTSFKVENLTGAIYYFRMAENEFVGVPVETVDVAKAABPPPPGKVTLTD 25150
Qy 134 ----- 133
Db 25151 VSQTSASLMWEKPEHGGSRVLGVVEMQPKOTEKMSIVAESKVCNAVVTGLSGQEQYOF 25210
Qy 134 ----- 133
Db 25211 RVKAYNEKGS DPRVLGVPVIAKDLTIQPSLKLPTNTYSIOAGEDLKIEIPVIGRPRNI 25270
Qy 134 ----- 133
Db 25271 SMVKDGEPLKQTRNVNVEETATSTVLHIKEGNKDDFGKTVTATNSAGTATENLSIVILE 25330
Qy 134 ----- 133
Db 25331 KCGPPVGPVRDEVSADFVVISWEPYPAYTGGCOISNYIVEKRDTTTTTHMVSAIVARTT 25390
Qy 134 ----- 133
Db 25391 IKITLKTGYQFRIFAENRYKGSAPLDSKAVIYOYPPKEPGPGPTPPTSISKDOMLV 25450
Qy 134 ----- 133
Db 25451 QMHEPVNDGGTKIIGYHLEQEKNSILWVKLNKTPIODTKFKTTGLDEGLEVEFKVSAEN 25510
Qy 134 ----- 133
Db 25511 IVGIGKPSKVSECFVARDPCDPPRPEAIVITRNNVTLKWKPAYDGGSKITGYIVEKKD 25570
Qy 134 ----- 133
Db 25571 LPDGRWMAKASFTNVLTEFTVSLVEDQRYEFVRVIAARNAAGNPFSPSDSQAITARDEID 25630
Qy 134 ----- 133
Db 25631 APNASLDPKYKDVIVVHAGETFVLEADIRGKPIPDVVMVSKDGELEBTAARMEIKSTIQK 25690
Qy 134 ----- 133
Db 25691 TTLVWKDCIRDTGGQYILKLSNVGKTSIPITVKVLDLDRPPEGLKVTGVTAEKCYLAW 25750
Qy 134 ----- 133
Db 25751 NPPLQDGGANISHYIIEKRETSRLSWTQVSTEVOALNYKVKLLPGNEYIFRVMVNVKYG 25810
Qy 134 ----- 133
Db 25811 IGEPLESGPVTAACNPKPPGPPSTPEVSAITKDSMVVTWARPVDDGGTEIEGYILEKRDK 25870
Qy 134 ----- 133
Db 25871 EGVWTKCNKKTLDLRLRVTLGTGHSYEFVRAAENAGVGPSESPVYRACDALYPP 25930
Qy 134 ----- 133

Db 25931 GPPSNPKVTDTSRSSVSLAWSKPIYDGGAPVKGYVVEVKEAAADEWTTCTPTTGLQKQF 25990
Qy 134 ----- 133
Db 25991 TVTKLKENTYFNFRICAINSEGUGEPATLPGSVQAQERIEPPEIELDADLRKVVVLASA 26050
Qy 134 ----- 133
Db 26051 TLRLFTVTKRPEPEVKEAEGILTDLRAQIEVTSFTMLVIDNVTRFDGRYNLTLENN 26110
Qy 134 ----- 133
Db 26111 SGSKTAFVNVRLDPSAPVNLTIREVKKDSVTLTSWEPPLIDGGAKITNYIIVEKBETTRK 26170
Qy 134 ----- 133
Db 26171 AYATINNCTKTFRIENLQEGCSYFRVLASNEYGIGLPAETTEPVKVSPEPLPPGRVT 26230
Qy 134 ----- 133
Db 26231 LVDVTRNTATIKWKPESDGGSKITGYVVMOTKGSEKWSCTQVKTLEATISGLTAGEE 26290
Qy 134 ----- 133
Db 26291 YVFRVAANEKGRSDPRLQCPVPIARDIEIKPSVELPFHTFNVKAREQLKIDVPFKGRPQ 26350
Qy 134 ----- 133
Db 26351 ATVNRKDGOTLXETTRVNVSSKTVTSLSIKEASKEDVGTLYELCVSNSAGSITVPITII 26410
Qy 134 ----- 133
Db 26411 VLDRPGPPGPIRIDEVSCSITISWNPPEYDGGCOISNYIIVEKETTSTWHIVSOAVAR 26470
Qy 134 ----- 133
Db 26471 TSIKIVRLTTGSEYQFRVCAENRYGKSSYESSAVNAEYFPSPGPGTKVHVHATKSTM 26530
Qy 134 ----- 133
Db 26531 LVTHQVPNDGSRVIGVHLEYKERSILWSKANKILIAOTOMKVSGLDEGLMYEYRVA 26590
Qy 134 ----- 133
Db 26591 ENIAGICKSKCEVPAPDCDPPGQPEVNTITRKSLSLWSPHYDGGAKITGIVER 26650
Qy 134 ----- 133
Db 26651 RELPDGRWLKCNVTNIQETFEVTELTEDORYEPRVFARNAADSVSPSESTGP11VKDD 26710
Qy 134 ----- 133
Db 26711 VEPFRVMDVKFRDVIIVKAGEVLKINADIAGRPLPVISWAKOGIEIEERARTEIISTDN 26770
Qy 134 ----- 133
Db 26771 HTLLTVKDCIRDTQGYVLT1KNVAGTRSAVNCVKLDKPPAGPAGLEINGLTAEKCSLS 26830
Qy 134 ----- 133
Db 26831 WCRPOEDGADIDYIIVEKRETSHLAWTICEGELQMTSKVTLLKGNIEYIFRVTGVNKY 26890
Qy 134 ----- 133
Db 26891 GVGEPLESVAIKALDPFTVPSPPTSLEITSVTKESMTLWCRPESDGGSEISGVIIERRE 26950
Qy 134 ----- 133
Db 26951 KNSLRVNRKPKVLDLRVKSTGLRGCEYEYRVVAENAGLSLPSETSPLIRAEDPVFL 27010
Qy 134 ----- 133
Db 27011 PSPPSKPIVDGKTTITIAWVKPLFDGGAPITGYTVBYKKSDDDTDWKTISIQLRGTEYT 27070

Qy 134 ----- 133
Db 27071 ISGLTTGAEEVFRVKSVNKVGASDPDSDDPQIAKEREBEPLFDIDSEMRKTLIVKAGAS 27130
Qy 134 ----- 133
Db 27131 FTMVPRGRPVNVLWSKPD0LRTRAYVDTTDSRTSLTIENANRDSOKYTLTIONVL 27190
Qy 134 ----- 133
Db 27191 SAASLT1LVVKVLDTPGPPPTNITVODVTKESAVLSWDVPENDGGAPVKNYHIEKREASKA 27250
Qy 134 ----- 133
Db 27251 WSVTNNCNRLSYKVTNLQEGAIYVFRVSGENEFVGIPAETKEGVKITEKSPSPKLG 27310
Qy 134 ----- 133
Db 27311 TSISKDSVSLTWLKPHEHGGSR1VHYVVEALEKGQKNWKCNAKSTHHVVVSGLRENEY 27370
Qy 134 ----- 133
Db 27371 FFRVFAENQAGLSDPRELLLPVLIKEQLEPPEIDMKNFPSSHVVYVVRAGSNLKVDPISGK 27430
Qy 134 ----- 133
Db 27431 PLPKVTLSDRGVPLKATMRFNTEITAENLTLINKESVTADAGRYEITAAANSSGTTKAFIN 27490
Qy 134 ----- 133
Db 27491 IIVLDRPGPTGPVVISDITEESVTLKWEPPKYDGGSOVTHYILLKRETSTAVMTVSAT 27550
Qy 134 ----- 133
Db 27551 VARTMKVMKLTGTGEEYQFRIKAENRFGISDHIDSACVTVKLPTVTPGPPSTPMTVNTVR 27610
Qy 134 ----- 133
Db 27611 ESITVGHWPVSGGSAVVGHYHLEMKDRNSILWOKANKLVIRTHFKVTTISAGLIYEFR 27670
Qy 134 ----- 133
Db 27671 VYAENAGVGKPSHPSEPVLAIDACEPPRNVRITDISKNSVLSWQQAFAFDGSGKITGYI 27730
Qy 134 ----- 133
Db 27731 VERRDLPGRWTKASFNTVETQFIISGLTQNSQYEFVFRVARNVAGSISNPSEVVGPIFC 27790
Qy 134 ----- 133
Db 27791 IDSYGGPVIDLPLEYEVVVKYRAGTSVKLRAGISGKPAPTIEWYKDDKELOTNALVCVEN 27850
Qy 134 ----- 133
Db 27851 TTDLASILIKDADRLNSGCYELKLRNMGASATIRVQILDKPPGPGPIEFKTVTABKI 27910
Qy 134 ----- 137
Db 27911 TLLWRPPADGGAKITHYIIVEKRETSRVVWVMVSEHLEECIIITTTKIIKGNIEYIFRVRV 27970
Qy 138 ----- 137
Db 27971 NKYIGIEPLESDSVAKNAFVTPGPGPIEVTKITKNSMTVWWSRPIADGGSDISGYFLE 28030
Qy 138 ----- 137
Db 28031 KRDKSLGFWKVLKETIRDTROKVTGLTENS DYQYRVCAVNAAGOGPFSEPFYKAADP 28090
Qy 138 ----- 137
Db 28091 IDPPGPAKIRIADSTKSSITLGMSPVYDGGSAVTGYVVEIROGBEEETVSTTSGEVR 28150

Qy 138 ----- 137
Db 28151 TTEVVSNLXPGVNYIFRVSANVCAGGEPENNEPVQAKDILEAPEIDLDVALRTSVIA 28210
Qy 138 ----- 137
Db 28211 KAGEDVQVLIPFKGRPPPTVTRKDEKNLGDARYSIENTDSSLLTIPOVTRNDTGKYI 28270
Qy 138 ----- 137
Db 28271 LTIENGVEPKSSTVSXKVLDTPAACQKLVKHVSRGTVTLLWDPPLIDGSPFIINXVIE 28330
Qy 138 ----- 137
Db 28331 KRDATKRTWSVUSHKCSSTSFQJDLSEKTPFFFRVLAENEIGIBCEPCTTEPVKAAEVP 28390
Qy 138 ----- 137
Db 28391 APIRDLNMDKSTKTSVILSWTKPFDGGSVITEYVWERKGBGTWGHAGISKTCEIEVS 28450
Qy 138 ----- 137
Db 28451 QLKESQVLEFRVFAKNEKGLSDPVTIGPITVKELIITPEVDLSDI PCAQVTVRIGHNVHL 28510
Qy 138 ----- 137
Db 28511 ELPYKGRPKPSISMLKDGLPLKESEFVRFSKTENKITLSIKNAKKEHGKYTVILONAVC 28570
Qy 138 ----- 137
Db 28571 RIAVPITVITLGPSPKPGPIRFDKADSVILSWDPEDNGGGEITCYSIKRETSQTN 28630
Qy 138 ----- 137
Db 28631 WKMVCSSVARTTFKVPNLVKDAEYQFVRAENRYGVSOPLVSSIIVAKHOFRIIPGPGKP 28690
Qy 138 ----- 137
Db 28691 VIYNTSDCMLTWDAPVYDGGSEVTGPHVEKERNILMOKVNTSPISGREYRATGLVE 28750
Qy 138 ----- 137
Db 28751 GLDYQFRVYAENSAGLSSPSPSKFTLAVSPVDPGTPDYIDVTRETITLKNPPLRDGG 28810
Qy 138 ----- 137
Db 28811 SKIVGSIEKROGNERWVRNFTDVSECYTGTGLSPGDRVEFPRIIARNAVGTISPSPQS 28870
Qy 138 ----- 137
Db 28871 SGIIMTRDENVPPIVEFGPEYFDGLI IKSGESLRICALVQGRPVPRVTFWKDGVIEIKRM 28930
Qy 138 ----- 137
Db 28931 NMEITDVLGSLFVRDATRDHRGVYTVKAKNAGSAAEIKVKQDTPGKVVGPIRTN 28990
Qy 138 ----- 137
Db 28991 ITGEKWTLMWADPLNDGCAPITHYIIEKRETSRLAWALIEDKCEAQSYTAIKLINGNEYQ 29050
Qy 138 ----- 137
Db 29051 FRVSANVKGVRPLSDPVAQIOYTVDPAPGIPSPNITGNSITLTWARPESDGGSEI 29110
Qy 138 ----- 137
Db 29111 QQVILERREKSTRVVKVISKRPISRETRFKVTGLTEGNEYEFHVMAENAGVGPASGISR 29170
Qy 138 ----- 137
Db 29171 LIKCREPVNPPGPTVVKVTDTSKTTVSLEWSPKVPDGGMEIIGYIEMCKADLGDWHKV 29230
Qy 138 ----- 137

Db 29231 NAEACVKTRYVTDLQAGEYKFRVSAINGAGKSDCEVTGTIKAVDRLTAPELDIDANF 29290
Qy 138 ----- 137
Db 29291 KQTHVVRAGASIRLFIAYQGRPTPTAVMSKPDNSLRAIDHITDTSFSTLTVCNCRNDA 29350
Qy 138 ----- 137
Db 29351 GKYTTLTVENNSGKSITFTVKVLDTPGPPGPIPTPKDVTGRSATLMDAPLDDGGARIHYY 29410
Qy 138 ----- 137
Db 29411 VVEKREASRRSHQVISEKCTROI FKNNDLAEGVPYFRVSAVNEVGCEPYEMPEPIVAT 29470
Qy 138 ----- 137
Db 29471 EQPAPPRLLDVVDTSKSSAVLAWLKPDDHGGSRITGYLLEMRKQSDFMVBAGHTKQJTF 29530
Qy 138 ----- 137
Db 29531 TVERLVEKTEYEFVRVAKNDAGYSEPREAFSVIIKEPOIEPTADLTGITNQLITCKAGS 29590
Qy 138 ----- 137
Db 29591 PFTIDVPISGRPAPKVTKLEMLKETDRVSIITTKDRTTLTVKDSMRGDSGRYFLTLE 29650
Qy 138 ----- 137
Db 29651 NTAGVKTFSTVVJIRPGPVTGPTEVSSVSAESCVLWGEKPGDGGTEITNIVIEKRES 29710
Qy 138 ----- 137
Db 29711 GTTAMOLVNSVVKRTOIKVTHLT KYMEYSFRVSSSENRFGVSKPLESAPIIAEHFPVPSA 29770
Qy 138 ----- 137
Db 29771 PTRPEVYHVSAANMSIRWEEPYHDGSKIYGVWEKERNTILWVENKVPCECNYKYVT 29830
Qy 138 ----- 137
Db 29831 GLVEGLEQFRTYALNAGVSKASEASRPMAONPVDAPGRPEVTDVTRSTVSLIWSAPA 29890
Qy 138 ----- 137
Db 29891 YDGGKVGVIIEERKPVSEVGDGRWLKCNVTIVSDNFTVTALSEGDTYEFVRLAKNAAG 29950
Qy 138 ----- 137
Db 29951 VISKGSSTGPVTCRDEYAPPAELDARLHGLVTIRAGSLVLDAAVGGKPEPKIIMTK 30010
Qy 138 ----- 137
Db 30011 GDKELDCEKVSLOYTGKRAVIAVKFCDRSDSGKYTLTVKNASGTVKASVNVKVLDSGP 30070
Qy 138 ----- 137
Db 30071 CGKLTVSRVTOEKCTLAWSLPQEDGCAETHYIVERRETSRLNWNVIVEGECPTLSYVWTR 30130
Qy 138 ----- 137
Db 30131 LIKNNEYIFRVAVKYGPVPESEPIVARNSTIPSPGIPPEVGTGKEHIIIOWTKP 30190
Qy 138 ----- 137
Db 30191 ESDGNEISNVLVDKREKSLRWRVKNKYVVDYDLKVTLSMEGCDYQFRVTVAVNAAGN 30250
Qy 138 ----- 137
Db 30251 SERSEANFISCREPSYTPGPSPAPRVVDTTKHSISLAWTKPMYDGTDIVGVYLEMOEK 30310
Qy 138 ----- 137

Db 30311 DTDQYRVHTNATIRNTEFTVPDLKXGQKYSFRVAANVKGMSSESIABIEPVERIEI 30370
Qy 138 ----- 137
Db 30371 PDLELADDLKKTVTIRAGASRLMVSGRPPPVITWSKQIDLASRAIIDTTESYSILLI 30430
Qy 138 ----- 137
Db 30431 VDKVNRDAGKYTIEAENQSGKSATVLVKVYDTFGPCSPSVKVEVSRDSVTITWEIPTI 30490
Qy 138 ----- 137
Db 30491 DGGAPNNYIVEKREAAAFKTVTTKCSKTLRISGLVEGTMYYFRVLPENIYIGIBPC 30550
Qy 138 ----- 137
Db 30551 ETSDAVLVSEVPLVPAKLEVDVTKSTVTLAMEKPLYDGGSLRTGYVLEACKAGTERWNK 30610
Qy 138 ----- 137
Db 30611 VVTLKPTVLEHTVTSNEGEQVLFRIARQNEKGVSEPRETVTAVTVQDLRVLPITDLSTM 30670
Qy 138 ----- 137
Db 30671 POKTHVPACRPVELVPIAGRPPPAASWFFPAGSKLRESERVTVETHTKVAKLTIRETTI 30730
Qy 138 EDTG----- 141
Db 30731 RDTGEVTELEKNVGTGTTSETIKVILDKPGPTGPIKIDEIDATISITWEPPELDGAP 30790
Qy 142 ----- 141
Db 30791 LSGYVVEQRAHRPGWL PVSESVTRSTFKPTRLTEGNEYVFRVAATNRFIGSYLQSEVI 30850
Qy 142 ----- 141
Db 30851 ECRSSIRIPGPETLQIFDVSRDGMTLWYPPEDDGSQVGYIVERKEVRADRWVRNK 30910
Qy 142 ----- 141
Db 30911 VPTMTRYRSTGLTEGLEVEHRTVAINARGSKPSRSPKPIVAMDPIAPPGKPNPRVTD 30970
Qy 142 ----- 141
Db 30971 TTRTSVLAWSVPEDEGSKVTGYLIEMQKVDQHEWTKNTTPTKIREYTLTHLPOGAEV 31030
Qy 142 ----- 141
Db 31031 RFRVLACNAGGPEAEVPGTVKVTMLEYDPDYELDERVQEGIFVRQGGVIRLTIPK 31090
Qy 142 ----- 141
Db 31091 PFPICKWTEGQDISKRAMIATSETHTELVIKEADRGDSGTVDLVLENKCGKKAIVIKVR 31150
Qy 142 ----- 141
Db 31151 VIGSPNSPEGLEVDYDIQVRSVRVSWRPPADGGADILGYILERREVPKAAWYVTDISRVR 31210
Qy 142 ----- 141
Db 31211 GTSLVWKGLENVEYHFRVSAENQFISKPLKSEEPVTPKTLNPPPPSNPPEVLDVTK 31270
Qy 142 ----- 141
Db 31271 SSVLSWSRPPKDDGSGRVTGYIERKETSTDKVVRHNKTQITTTMYTVTGLVDAEYQFR 31330
Qy 142 ----- 141
Db 31331 IIAQNDVCLSETSPASEPVCKDFDKPSQPGLELILSKDSVTLQWKEPCDGGKEIL 31390
Qy 142 ----- 141
Db 31391 GYWVEYRQSGDSAWKSKNERIKDKQFTIGLLEATEYFRVFAENETGLSRPRRTAWSI 31450

Qy 142 ----- 141
Db 31451 KTKLTSGEAPIRKEMKDVTTKLGEAAQLSCQIVGRPLDKWRFKELIOSRKYMSS 31510
Qy 142 ----- 141
Db 31511 DGRTHTLVTMTBEQDEGVYTCIATNEVGEVETSSKLLQATQFPHGYPKKEKYGAVG 31570
Qy 142 ----- 141
Db 31571 STLRLHVMYIGRPVPAMTWFGOKLLQNSENIENTEHTYHLVMKNVORKTHAGKYKV 31630
Qy 142 ----- 141
Db 31631 LSNVFGTVDAILDVEIQDKPKTGPVIEBALLKNSAVISWKPPADGGSMITNYVVVEK 31690
Qy 142 ----- 141
Db 31691 EAKEGAEMQLVSSAISVTTCTRIVNLTENAGYFRVSAQNTFGISDPLEVSVVLIKSPFE 31750
Qy 142 ----- 141
Db 31751 KPGAPGPTTAVTKDSCVVAWKPAPSDGGAKIRNYYLEKREKKONKWSVTTTEIRETV 31810
Qy 142 ----- 141
Db 31811 FSVKNLIEGLEVEPRVKENLGSEWSEISEPITPKSDVPIOAPHKPEELRNLNRYQS 31870
Qy 142 ----- 141
Db 31871 NATLVCKVTGHPKPIVKRYRGKEIIADGLKYRIOEFGYHQLIIASVTDVDDATVQVR 31930
Qy 142 ----- 141
Db 31931 ATNOGGSVGTASLEVEVPAKIHLPKTLEGMGAVHALRGEVVSIIKIPSGKPDVPITWQK 31990
Qy 142 ----- 141
Db 31991 QODLIDNNGHYQVIVTRSFSLVPFNGVERKDAGFYVCAKVRFGIDOKTVELDVADVPD 32050
Qy 142 ----- 141
Db 32051 PPRGVKVSVDVSRDSVNLTWTEPASDGGSKIINYIVEKCATTAERWLRVGOARETRYVIN 32110
Qy 142 ----- 141
Db 32111 LFGKTSYQFRVIAENKFGLSKSESEPTITKEDKTRAMNYDEEVDRETSMTKASHSS 32170
Qy 142 ----- 141
Db 32171 TKELYEKYMIABDLGRGFGIVHRCVETS SKKTYNAKFKVKVKGTDQVLVKKEISILNIAR 32230
Qy 142 ----- 141
Db 32231 HRNIIHLHESPESEMEELVMIPEFISGLDIPERINTSAFELNREIVSVYHVCEALQFLH 32290
Qy 142 FDL----- 144
Db 32291 SHNIGHFDIRPENIIVQTRRSSTIKIIBFGOARQLKPGDNFLLFTAPEYAPAEVHQHVD 32350
Qy 145 ----- 144
Db 32351 VSTATDMWSLGLTVVLLSGINPFLAETNQIIENIMNAEYTFDEAPKFEISIEAMD FVD 32410
Qy 145 ----- 144
Db 32411 RLLVKERKSRMTASEALQHPWLKQKIERVSTKVIRTLKHRRYYHTLIKKDLNMVWSAARI 32470
Qy 145 ----- 144
Db 32471 SCGATRSQKGVSAKVAKVASIEIGFVSGQIMHAVGEGGHVKYCKIENYDOSTQVTWY 32530

QY 59 ----- 58
 Db 2770 VSVSHDTVPVKNPHKSVKIPKSDKRLVSRKVHKLMLQNISPSDAGEYTAVVGOLECKA 2829
 QY 59 ----- 58
 Db 2830 KLFVETLHITKMTNIEVPETKTASFECEVSHFNVPFMMKNGVEIEMSEKFIIVVOGKL 2889
 QY 59 ----- 58
 Db 2890 HOLIIMNTSTEDSAEYTFVCGNDQVSATLTVTPIMITSMKLDINASEKOTITFEVTVNVE 2949
 QY 59 ----- 58
 Db 2950 GISYKWLKNGVEIKSTDKQWRCKKLTHSLNIRNVHFGDAADYTFVAGKATSTATLYVEA 3009
 QY 59 ----- 58
 Db 3010 RHIEFRKHIKIDIKVLEKKRAMFECEVSEPDITVOMMKDDOELQITDRIKIOKEYVHRL 3069
 QY 59 ----- 58
 Db 3070 IPSTRMSDAGKYTVVAGNVSTAKLFVEGRDVRIRSIKKEVOVIEKORAVVEFEVNEDDV 3129
 QY 59 ----- 58
 Db 3130 DAHWYKDGIEINFQOERHKYVVERRIHRMFISETROSDAGEYTFVAGRNRSSVTLVNA 3189
 QY 59 ----- 58
 Db 3190 PEPPOVLOEQPVTVQSGKPARFCAMISGRPOPKISWYKEQLLSTGFKFLHDGQEYT 3249
 QY 59 ----- 58
 Db 3250 LLLIEAFPDAAYTCEAKNDYGVATTSASLSVEVPEVSPDQEMPVYPPIITPLQDTV 3309
 QY 59 ----- 58
 Db 3310 TSEGQPARFOCRVSGTDLKVSWSXDKKIKPSRFRMTQFEDTVOLEIAEAYPEDEGTYT 3369
 QY 59 ----- 58
 Db 3370 FVANNAGOVSTANLSLEAPESILHERIEQIEIEMEMKEFSSFLSABEEGLHSAELQLS 3429
 QY 59 ----- 58
 Db 3430 KINETLELLSESVPYTKFDESEKGTGPIFIKEVSNADISMGDVATLSVTWIGIPKPKIQ 3489
 QY 59 ----- 58
 Db 3490 WFFNGVLLTPSADYKFVFDGDDHSLIILFTKLEDEGEYTCMASNDYGKTCISAYLKINSK 3549
 QY 59 ----- 58
 Db 3550 GEGHKDTETESAVAKSLEKLGPCPPHFLKELKPIRCAOGLPAIPEYTVVGEPAPTVMF 3609
 QY 59 ----- 58
 Db 3610 KENKQLCTSVYTYIHNPNNGSGTFIVNDPQREDSGLYICKAENMLGESTCAAELLVLED 3669
 QY 59 ----- 58
 Db 3670 TDMTDPCKAKSTPEAPDPOTPLKGPAAVEALDSEQEIATFVKDTILKAALITEENQOL 3729
 QY 59 ----- 58
 Db 3730 SYEHIKANELSQLPLGAQELQILQDKLTPESTREFLCINGSIHFPQPKPEPSNLQL 3789
 QY 59 ----- ILTEPEG- 65
 Db 3790 QIVOSQKTFSEKILMPEPETOAVLSDTEKIFPSAMSIEQINSLTVEPLKTLAEPEGN 3849
 QY 66 ----- 65

Db 3850 YPOSSIPEPPMHSYLTSAEEVLSUKKTVSDTNREORVTLOKQOASQALILSOSLAEGHV 3909
 QY 66 ----- 65
 Db 3910 ESLOSQPDVMIQVNYEPLVPSEHSTEGGKILIESANPLENAGQDSAVRIEKGSLRPL 3969
 QY 66 ----- 65
 Db 3970 ALEEKQVLLKEEHSNDVMPDQIIIESKREPVAIKKQVQVGRDLLSKESLLSGIPERQ 4029
 QY 66 ----- 65
 Db 4030 LNLKTIQICRALQAAVASQPGLFSEWLRNIEKVEAVEAVNIQEPHRHIMCMVLYTSKSVT 4089
 QY 66 ----- 65
 Db 4090 BEVTIIIEDVDPMANLKMELRDALCAIYBEIDILTAEGPRIQOGAKTSLOEENDSFSG 4149
 QY 66 ----- 65
 Db 4150 SOKVEPITEPEVESKYLISTEEVSYFNQSRVKYLDATPVTKGVASAVVSEKODESLPK 4209
 QY 66 ----- 65
 Db 4210 SEEKESSSESCTEEVATVKIQEAEGLIKEDGPMIHTPLVDTVSEGDIVHLTTSITNA 4269
 QY 66 ----- 65
 Db 4270 KEVNYFENKLVPSDEKFKCLOQONTYTLVIDKNTEDHOGVEVCEALNDSKGTATSACL 4329
 QY 66 ----- APNI----- 69
 Db 4330 TVVKRAAPVTKRIEPLVALGHLAKFTCEIQSAPNVRPQWPKAGREIYESDKCSIRSSK 4389
 QY 70 ----- 69
 Db 4390 YISSLEILRTQVVDCGEYTKASNEYGSVSTATLTVTPGGEKKVKRLLPERKPEPKE 4449
 QY 70 ----- 69
 Db 4450 VVLKSVLRKRPESEEPKVPKLEKVKPAVPEPPPPKPEVEVEVPTVKRERKIPEPTK 4509
 QY 70 ----- 69
 Db 4510 VPEIKPAIPLPAPEPKPAEAVKTIKPPPEPEPTPIAAPVTVPVVVGKKAAPKEEA 4569
 QY 70 ----- 69
 Db 4570 AKPGPIKGVPKTPSPIEARRKLRPGSGGKPPDEAPPTYQLKAVPLKPVKEIKDIL 4629
 QY 70 ----- 69
 Db 4630 TESEFVGSSAIFECVLSFSTAITTMMKDGSIRESKPHRFIADGDKRKLHIIDVQLSDAG 4689
 QY 70 ----- 69
 Db 4650 EYTCVLRGNKEKSTAKLVVEELPVRFVKLBEETVVKGPPLYLSCELNKERDVMRK 4749
 QY 70 ----- 69
 Db 4750 DGKIWEKPGRIVPGVIGLMRALTINDADDTAGTYTVTVENANNLESCSCVKVVEIRD 4809
 QY 70 ----- 69
 Db 4810 WLVPKIRDOHVKPGTAIFACDIAKDTPNIKWFKGYDEIPAEPNDKTILRDGNHLYUKI 4869
 QY 70 ----- 69
 Db 4870 KNAMPEDIAEYAVEIEGKRYPAKTLGERVELLKPDIETVIEKESASPDALISEADIP 4929
 QY 70 ----- 69

Db	4930	QWKLKGELLRSPPTCEIKAEGRFLTLHKVLQDAGEVLYQALNAITTAITVKEIEL	4989
Qy	70	-----	69
Db	4990	DFAVPLKDVTPPERQARFECVLTREANVWSKGPDIISKSDKFDIIADGKHILVINDS	5049
Qy	70	-----	69
Db	5050	QFDBGVVTAEEVGKTSARLFTVGTIRLKFMSPLEDQTVKEGETATFVCELSHEKQWVW	5109
Qy	70	-----	69
Db	5110	FKNDAKLHSTRVLISSEKTHLEMKVETLDDISQIKAQVKELSSSTAQLKVLEADPYFT	5169
Qy	70	-----	69
Db	5170	VKLHDKTAVEKDEITLKCEVSKDVPKWFKDGEETVPSKYSIKADGLRRIKIKKADLK	5229
Qy	70	-----	69
Db	5230	DKGEVVCDCGTDKTKANVTVEARLIEVEKPLYGVEFVGETAHEIELSEPDVHGQWKLK	5289
Qy	70	-----	69
Db	5290	GOPLTASPDCEIIEGKKHILIHNCQLGWTGEVSFOAANAKSAANLKVYKELPLIFITPL	5349
Qy	70	-----	69
Db	5350	SDVKVFEKDEAKFECEVSREPFTFRWLKGTQBITGDRFELIKDGTQKSHWIKSAAFEDE	5409
Qy	70	-----	69
Db	5410	AKYMFEAEDKHTSGKLIIEGIRLKLTPDKVTAKEKESAVFTVELSHDNIRVWFKNDQ	5469
Qy	70	-----	69
Db	5470	RLHTTRSVSQDEGKTHSITFKDLSIDDTSQIRVEAMGMSSEAKLTVLEGGDPYFTGKLQD	5529
Qy	70	-----	69
Db	5530	YTGVEKDEVILQCEISKADAPKMFKDGEIKPKSKNAVIKTDGKXRMILILKALKSDIQQ	5589
Qy	70	-----	69
Db	5590	YTCDCGTDKTSGLKDIEDREIKLVRPLHSVEMETETARFETEISEDDIHANWKLGEAL	5649
Qy	70	-----	69
Db	5650	LQTPDCEIKEEGKIHSVLHNCRLDQGGVDFQAAVVKSSAHLRVKPRVIGLLRPLKDVY	5709
Qy	70	-----	69
Db	5710	VTAGETATFDCELSYEDIPVEWYLGKKLEPSDKVVPRSEGVHTLTLRDVKLEDAGEVQ	5769
Qy	70	-----	69
Db	5770	LTAKQFKTHANLKVKEPPVEFTKPLEDQTVKEGATAVECEVSRENKVKWFKNGTEILK	5829
Qy	70	-----	69
Db	5830	SKKYEIVADGRVKLVHDCTPEDIKTYTCDKDFKTSNCLNVVPPHVEFLRPLTDLQVR	5889
Qy	70	-----	69
Db	5890	EKEMARFECELSRENKVKWFKDGAIEIKGKYDIISKGAVRILVINKCLLDDAEAYSCE	5949
Qy	70	-----	69
Db	5950	VRTARTSGMLTVLEEBAVFTNLANIEVSETDTIKLVCEVSKPQAEVWYKGDDEIIEGT	6009
Qy	70	-----	69
Db	6010	RYEILTGRKRILVIONAHLEDAGNVCNRLPSSRTDGVKVKVHELAAEFISKQNLFILEG	6069
Qy	70	-----	69
Db	6070	EKAFFVCSISKESFPVQWKRDDKTLESQKDYDVIADGKKRVLVVKDQATLQDMCTYVVMVG	6129
Qy	70	-----	69
Db	6130	AARAAHLTVIEKLRIVVPLKOTRVKEQOEVEVFNCEVNTGAKAKWFRNEAIFDSKYI	6189
Qy	70	-----	69
Db	6190	ILOKDLVYTLIRDAHLDDQANVNSLTNHRGENVKSAAANLIVEEEDLRIVEPLKDIETM	6249
Qy	70	-----	69
Db	6250	EKKSVTFCWCKVNLNVLTKWTKNGEEVVFONRVSYRYDKYKHLMTIKDCGFPDGEYIVT	6309
Qy	70	-----	69
Db	6310	AGQKSVAEILLIIEAPTEFVEHLEDQTVTEFDDAVFSCQSREKANVKWYRNGREIKEGK	6369
Qy	70	-----	69
Db	6370	KYFEKDGSIHRLIIEKDLDDCEYACGVEDRKSRLRFLVEEIPVEIIRPPQDILEAPG	6429
Qy	70	-----	69
Db	6430	ADVFLAELNKDKVEVQWLRNNMVVQGDKHQWSEKGIHRLQICDIKPRDQGEYRPIAK	6489
Qy	70	-----	69
Db	6490	DKEARAKLEAAAPKIKTADQDLVVDVGKPLTMVVPYDAYPKABAEWFKENEPLSTKIID	6549
Qy	70	-----	69
Db	6550	TTAOTSFRILBAKKGDKGRYKIVLQNHGKABGFINLKVDPVGRNLEVTTETPDGEV	6609
Qy	70	-----	69
Db	6610	SLAMEEPLTDGSKIIGVWVERRDIKRKTWVLATDRAESCEFTVTGLQKGVEYLFVSA	6669
Qy	70	-----	69
Db	6670	RNRVGTGEVETDNPVEARSKYDVGPPPLNVTTIDVNRFGVSLTWEPPEYDGGAEITNV	6729
Qy	70	-----	69
Db	6730	IELRDKTSIRWDATVRAEDLSATVTDVWEGQEYSFRVRAQNRIGVQKPSAATPFVKVA	6789
Qy	70	TANLTSS-----	76
Db	6790	DPIERPPVNLTSQDQSSVQLKWBPPPLKDGGSPILYIIERCEGKNWIRCNMKLV	6849
Qy	77	-----	76
Db	6850	PELTYKVTGLEKGNKLYRVAENKAGVSDPSEILGLPLTADDAFVEPTMDLSAFKDGLEV	6909
Qy	77	-----	76
Db	6910	IVPNPITILVPSTGYPRPTATMCFGDKVLETDGRVQMKTL SAYAELVISPSERSDKIYT	6969
Qy	77	-----	76
Db	6970	LKLENRVKTSIGEIDVNVIAIARPSAPKELKFGDITKDSVHLTWEPDDDDGGSPLTGYVVEK	7029
Qy	77	-----LLSVC-----	81
Db	7030	REVSRTKTKVMDFTDLEFTVVDLVQGEYLFVKVCAENKCGPGEPAVDEPNVNSTPAT	7089
Qy	82	-----	81
Db	7090	VPDPENVKWRDRTANSIFLTWDPKNDGGSRIKGYIVERCPRGSKWVACGEPVAETKM	7149

Qy	82	-----	81
Db	7150	EVTGLEBKWYAVRVKTLNRQASKPSRPTETIEIQAVDTQEAPEIFLDVKLLAGLTVKACT	7209
Qy	82	-----	81
Db	7210	KIELPATVTVGKPEPKITWTAKDMILKQDKRITITENVPKSTVTIIVDSKESDGTGTIIEAV	7269
Qy	82	-----	81
Db	7270	NVCGRATAVVENVLDPGPPAAFDITDVNESCLLTWNPPRDDGGSKITVYVVERRATD	7329
Qy	82	-----	81
Db	7330	SEVHKLSTVKOTNFKATKLIIPNKEYIFRVAENMYGAGEPVQASPIITAKYQFDPGPP	7389
Qy	82	-----	81
Db	7390	TRLEPSDITKDAVTLTWCEPDDGGSPITGYWVERLDPDTDKWRCNKMVPKDTTYRVKG	7449
Qy	82	-----	81
Db	7450	LTNKKYFRVRLAENLAGCGKPSKSTEPILIKDIPDPMPGKPTVKDVGKTSVRLNWTK	7509
Qy	82	-----	81
Db	7510	PEHGGAKIESYVIEMLKTCTDEVRVAEGVPTTOHLLPGLMEGOEYSFRVRAVNAKAGES	7569
Qy	82	-----	81
Db	7570	EPSEPSDVLCKREKLYPPSPRWLEVINITKNTADLKWTVPEKDGGSPITNIVIEKRDVR	7629
Qy	82	-----	81
Db	7630	RKGHQTDDTVTKTKCTVTLTEGLYVFRVAENAIQSDSYTEIEDSVLAKDTFTTPGP	7689
Qy	82	-----	81
Db	7690	PYALAVVDVTKRHVDLKWEPKNDGGRPIORYVIEKKERLGRVWVAKGTAGDCNFRVT	7749
Qy	82	-----	81
Db	7750	DVIEGTEVOFVRAENEAUGHPSPTTEILSIEDTSPSPPLDLHVTDAGRKHIAIAWK	7809
Qy	82	-----	81
Db	7810	PPEKNGSGPIIGYHVEMCPVTEKMRVNSRPIKDLKFVKEGVVPDKEYVLRVAVNAI	7869
Qy	82	-----	81
Db	7870	GVSEFSEISENVAKDPCKPTIDLETHDIIVIEGEKLSIPVPPRAVPVPTVSHKDGKE	7929
Qy	82	-----	81
Db	7930	VKASDRLTMKNDHISAHLEVPKSVRADAGIYVITITLENKLSATASINVKVLGPCKDI	7989
Qy	82	-----	81
Db	7990	KASDITKSSCKLTWEPPEFDGGTPILVYLERREAGRTYIPVMSGENKLSWTVKDLIPN	8049
Qy	82	-----	81
Db	8050	GEYFPRVAVNKGGEYIELKNPVIADPKQPPDPVDEVHNPTAEAMTITWKPLDYD	8109
Qy	82	-----	81
Db	8110	CGSKIMGYIEKIAKEERWRCNEHLVPILTYTAKLEGKEYQFVRVAENAAAGISEPS	8169
Qy	82	-----	81
Db	8170	RATPPTKAVDPIDAPKVILRTSLEVKGDEIALDASISGSPYPTITWIKDENVIIVPEIK	8229
Qy	82	-----	81

Db	8230	KRAAPLVRRRKGEVQEBEPFVLPLTQRLSIDNSKKGESQSLRVDRSLRPHGLYMIKVEND	8289
Qy	82	-----	81
Db	8290	HGIAKAPCTVSVLDTPOPPINFVPEDIRKTSVLCKWBPPLDDGGSEIINYITLEKKDKTKP	8349
Qy	82	-----	81
Db	8350	DSEWIVVTSTLRHCKYSVTKLIEGKEYLFRVRAENRFGPGPCVSKPLVAKDPFGPPDAP	8409
Qy	82	-----	81
Db	8410	DKPIVEDVTSNMLVKWNEPKDNGSPILGYWLEKREXVNSTHWSRVNKSLLNALKANVDGL	8469
Qy	82	-----	81
Db	8470	LEGLTYFRVCAENAAAGKFSPPSDPKTAHDPIPPOPPIPRVTDTSSTTIELEWEPPA	8529
Qy	82	-----	81
Db	8530	FNGGGEIVGYFVDKQLVGTNKNKSRCTEKMIKVROYTVKEIREGADYKLRVSAVNAAGEP	8589
Qy	82	-----	81
Db	8590	PGETQPTVABPQEPPEPAVELDVSVKGGTQIMAGKTLRIPAVVTCRVPVTKVMTKEEGEL	8649
Qy	82	-----	81
Db	8650	KORVVVIDNVGTSKELIIKDALRKDHGRYVITATNCSGSKFAAARVEFVDPGVLDLKPV	8709
Qy	82	-----	81
Db	8710	VTNRKWCLLNMSDPEDDGGSEITGFIIERKDAKMHTRQPIETERSKCDITGLLEGQBYK	8769
Qy	82	-----	81
Db	8770	FRVIAKNKFGCGPPVEIGPILAVDPLGPPTSPELTYTERORSTITLOWKEPRSGGSP	8829
Qy	82	-----	81
Db	8830	QGYIEKRHRDKPDFERVNKRCLPTTSFLVENLDEHOMYFRVKAIVEGESEPSLPLNV	8889
Qy	82	-----	81
Db	8890	VIQDDEVPPTIKLRLSVRGDTIKVKAGEPVHIPADVTGLPMPKIEWSKNETVIEKPTDAL	8949
Qy	82	-----	81
Db	8950	QITKEVSRSEAKTELISIPKAVREDKGTVTVTASNRLGSVFRNVHVEVYDQSPPRNLAV	9009
Qy	82	-----	81
Db	9010	TDIAKESCYLTWDAPLDNGGSEITHYVIDKRDASRKKAEWEVTNTAVEKRYGIMKLIPN	9069
Qy	82	-----	81
Db	9070	GQYEFVRVAVNKYGISDECKSDKVIQDPYRLPGPPGPKVLARTKGSMLVSWTPPLDNG	9129
Qy	82	-----	81
Db	9130	GSPITGYWLEKREEGSPYWSRVRAPITKVGLKGVEFNVPRLLLEGVKYQFRAMAINNAGI	9189
Qy	82	-----	81
Db	9190	GPFSESDPEVAGDPIFPPOPPSPCEVKDKTKSSIISLGWPKPAKGGSPKIGYIVEMQEE	9249
Qy	84	-----	83
Db	9250	GTTDKWRVNEPDKLITTCCEVWPNLKLRRYFRVKAIVEGESEPSDITTEIPATDIOE	9309
Qy	84	-----	83

Db 9310 EPEVFIDIGAQCCLVKAGSQIRIPAVIKGRPTPKSSWEFDGKAKKMGVHDIPEDAQ 9369
Qy 84 ----- 83
Db 9370 LETAENSSVIIPECKRSHTGYSITAKNKAQKQKANCVRKVMVDPGPKOLKVSIDIIRG 9429
Qy 84 ----- 83
Db 9430 SCRLSKWKPDDGGDRIKGYVIEKRTIDGKAWTKYNPCDGSFTFVVPDLLSEQQYFFVRV 9489
Qy 84 ----- 83
Db 9490 AENRFGIGPPVETIORTTARDPIYPPDPPIKLGITKNTVHLSWKPPKNDGGSPTVHY 9549
Qy 84 ----- 83
Db 9550 IVECLAWDPTGKKEAMROCNKRDEELOFTVEDLVEGGEYEFVRKAVNAAGVSKPSATV 9609
Qy 84 ----- 83
Db 9610 GPDCORPDPSPSIDLKEFMEVEEGTNVNI VAKIKGVPPTLTWFKAPPKPDNKEPVLV 9669
Qy 84 ----- 83
Db 9670 DTHVKNLVDDTCTLVIPOSRRSDTGLYITITAVNNLGTASKEMRLNLVGRPGPVGPIKF 9729
Qy 84 ----- 83
Db 9730 ESVSADQMTLSWFPKDDGSKIITNYVIEKREANRKTWVHVSSEPKECTYITIPKLEGE 9789
Qy 84 -----SQTIN----- 88
Db 9790 YVFRIMAQNKYGIGELDSEPETARNLFSVPGAPDKPTVSSVTRNSMTVNWEEPEYDGS 9849
Qy 89 ----- 88
Db 9850 PVTGYWLEMKDTSKRWRVNRDPIKAMTLGVSYKVTGLIEGSDYQFRVYAINAAGVGA 9909
Qy 89 ----- 88
Db 9910 SLPSDPATARDPIAPGPPPKVTDWTKSADLEWSPPLKDGSKVTGYIYEKGEKKE 9969
Qy 89 ----- 88
Db 9970 WEKGDKEVRGTLVVTGLKEGAFYKFRVSAVNIAGIGEPGEVTDVIEMKDLVSPDLQL 10029
Qy 89 ----- 88
Db 10030 DASVRDRIVVHAGGVIRIIAYVSGKPPPTVTNMNERTLPQEATIIETAISSSMVKNCO 10089
Qy 89 -----PEDD--- 92
Db 10090 RSHGVVYLLAKNEAGERKKTIIIVDVLDPGVGTPTFLAHLNLTNESCCLTWFSPEDDGGS 10149
Qy 93 ----- 92
Db 10150 PITNYVIEKRESDRATPVTYVTRONATVQGLIQGKAYFFRIAAENSIGMPFVETSE 10209
Qy 93 ----- 92
Db 10210 ALVIREPITVPERPEDLEVKEVTNVTLTWNPYPKYDGGSEIINYVLESRLICTEKFHKV 10269
Qy 93 ----- 92
Db 10270 TNDNLSRKVTYKLGKEGDTYEVRSVAVNIVGQKPSFCTKPITCKDELAPPTLHLDFRD 10329
Qy 93 ----- 92
Db 10330 KLTIIRGEAFALTGRYSKPKPKVSWFKDEADVLEDDRTHIKTTPATLALEKIKAKRSDS 10389
Qy 93 ----- 92
Db 10390 GKVCVVVENSTGRKGFQCNVVDHPGPPVGPVSFDEVTKDYMWVISMKPPLDDGGSKITN 10449

Qy 93 ----- 92
Db 10450 YIIEKKEVGKDVMPVPTSASAKTTCKVSKLEGGDYIFRIHAENLYGISDPLVSDSMKAK 10509
Qy 93 ----- 92
Db 10510 DRFRVPADPOPIVTEVTKDSALVTWNKPHDGGKPIITNYILEKRETMKRWARVTKDPIH 10569
Qy 93 ----- 92
Db 10570 PYTKFRVPDLLLEGQCQYBFRVSAENEIGIGDPSBSPKVPFAKPIAKPSPPVNPDAIDTTC 10629
Qy 93 ----- 92
Db 10630 NSVDLTWQPRHDGGSKILGYIVSYKVQGEEMRRANHTPESCPETKYKVTGLRDGQTYK 10689
Qy 93 -----TDPGH----- 97
Db 10690 FRVLAVNAAGESDPAHVPEPVLVKDRLEPPPELILDANWAREQHIKVGDTLRLSAILKGVP 10749
Qy 98 ----- 97
Db 10750 FPKVTWKEDRDAPTKARIDVTPVGSKLEIRNAAHEDGGIYSLTVENPAGSKTVSVKVLV 10809
Qy 98 ----- 97
Db 10810 LDKPGPRDLSEVBIRKDSCLTWEPLDDGGSVITNVVERRDVAQAQMSPLSATSXXX 10869
Qy 98 ----- 97
Db 10870 SHFAHLNEGQYLFRVAANQYGRGFVETPKPIKALDPLHPGPKDLHHVDVDKTEV 10929
Qy 98 ----- 97
Db 10930 SLVWNKPDGGSPIITGYLVEYQSEGTQDMIKFKVTNLECVVTLQOGKTYRFRVKAEN 10989
Qy 98 ----- 97
Db 10990 IVGLPLPTTIPIECQEKLVPPSVELDVKLIIEGLVVKAGTTVRFPPIIRGVVPVPTAKWTT 11049
Qy 98 ----- 97
Db 11050 DGSEIKTDEHYTETDNFSSVLTIKNCLRRDTGEOITVSNAGSKTVAVHLTVLDVPGP 11109
Qy 98 ----- 97
Db 11110 PTGPIINILDVTPHEMTISWQPKDDGSPVINIYVEKODTRKDTMGVVSSGSKTKLKIP 11169
Qy 98 ----- 97
Db 11170 HLQKGEYVFRVRAENKIGVGPLDSTPTVAKHFPSPSPGPKPVVTDITENAAATVSWTL 11229
Qy 98 ----- 97
Db 11230 PKSDGSPITGYMERREVTGKWRVKNKTPADLKFRVTGLYEGNTYEFVRFAENLAGLS 11289
Qy 98 -----ADLV----- 101
Db 11290 KPSPSDDPIKACRPIKPPGPPINPKLKDKSRETADLVWTKPLSDGGSPILYGVVECOKPG 11349
Qy 102 ----- 101
Db 11350 TAQWNRINKDELIROCAFVRVPGVGLIEGNEYRFRKAANIVGEGERELAESVIAKDLHHP 11409
Qy 102 ----- 101
Db 11410 EVELDVTCDVITVRVQGTIRILARVKGPEPDIITWTKEGKVLVREKRVLDLIQDLPRVEL 11469
Qy 102 ----- 101
Db 11470 QIKEAVRADHGKVIISAKNSSGHAQSAIVNVLDPRGPCONLKVNTVTKENCTIISWENPL 11529

QY	102	-----	101
Db	11530	DNGSEITNFIVEYRKPQKQMSIVASDVTKRLIKANLANNEYFRVCAENKVGVP	11589
QY	102	-----	101
Db	11590	ETKTPILAINPIDRGPENLHIADKGTFFYLVKWRPDDGGSPNLSYHVERLLKGSDD	11649
QY	102	-----	101
Db	11650	WERVHKSIGETHYMDRCVENQIYEFRTVOTKNEGSEDMVKTBEVVVKEDLOKPVLDLK	11709
QY	102	-----	101
Db	11710	LSGVLTVKAGDTIRLEAGVRGQPEVAWTKDADTLTRSPRVKIDTRADSSKFSLTKA	11769
QY	102	-----	101
Db	11770	KRSDGGKVVVTTATNTAGSFVAYATVNVLDKPGVPRNLKIVDVSSDRCTVCWDPDDGGC	11829
QY	102	-----	101
Db	11830	EIQNVILEKCEKRMVWSTYSATVLTGTTVTRLIEGNEVIFRVAENKIGTGPPTESKP	11889
QY	102	-----	101
Db	11890	VIAKTYDKPGRPPPEVTKVSKEMTVMWNPPEYDGGKSIPTYFLEKKEKHSTRVVPVN	11949
QY	102	-----	101
Db	11950	KSAIPERBMKVONLLPDHEYQFRVKAENEIGIGEPSLPSRPVAKDPIEPGPTNFRVV	12009
QY	102	-----	101
Db	12010	DTTKHSITLWGKPVYDGGAPIIGYVEMRPKIADASDEGWKRCNAAQLVRKEFTVTS	12069
QY	102	-----	101
Db	12070	LDENQYEFVCAQNVGIGRPAELKEAIKPEILEPPEIDLDASMRKLIVRAGCCIRL	12129
QY	102	-----	101
Db	12130	FAIVGRPAPKVTRKVGIDNVVRKQVDLVDTWAFVIPNSTRDDSGKSYLTILVNPAGE	12189
QY	102	-----	101
Db	12190	KAVFVNRVLTGPGVSDLVKTSCHVSWAPPENDGGSQVTHYIVEKREADRTWS	12249
QY	102	-----	101
Db	12250	TVTPEVKKTSFHVNLVPGNEYFRVTAINEYGPVPTDVPKPVLASDPLSEDPDPKLE	12309
QY	102	-----	101
Db	12310	ATEMTKNSATLAWLPLRDGAKIDGYIISYREEEQPADRWTEYSWKDLSLVWTGLKEG	12369
QY	102	-----	101
Db	12370	KKYKFRVAARNAVGVSLPREAGVYEAKELPPKILMPEQITIKAKKLRIEAHVYKGP	12429
QY	102	-----	101
Db	12430	HPTCKWKGEDEVTTSSHLAVHKADSSSILIIKDVTRKDSGYSLTAENSSGTDQKIKV	12489
QY	102	-----	101
Db	12490	VWMDAPGPQPPFDISIDADACSLSWHIPLEGGSNITNYIVEKCDVSRGDMVTALASV	12549
QY	102	-----	101
Db	12550	TKTSCRVGKLIHQEYIFRVAENRFGISEPLTSPQWVAQFPFVPEPKNARVTKVND	12609
QY	102	-----	101

Db	12610	CIFVANDRPDSGGSPIIIGYLIERKERNLLWVKANDTLVRSTEYPCAGLVEGLESPRI	12669
QY	102	-----	101
Db	12670	YALNAGSPSPKPTTEYVYARMFVDPGKPEVIDVTKVSLIWARPKHGGSKIIGYFV	12729
QY	102	-----	101
Db	12730	EACKLPDGKWRVCNTAPHQIQOEYATATCLEEKAQYQFRAIARTAVNISPPSEPDPVTI	12789
QY	102	-----	101
Db	12790	LAENVPRIIDLSVAMKSLTLVKAGTNVCLDATVFGKPMPTVSWKKDGTLLKPAEGIKWAM	12849
QY	102	-----	101
Db	12850	QRNLCTLEFSVNRKDSGDYTTIATENSSGSKSATIKLVLDKPGPPASVKINKMYSDRAM	12909
QY	102	-----	101
Db	12910	LSWEPLEDGGSEITNYIVDKRETSRPNQAVSATVPITSCSVEKLEHGEHYOFRICAEN	12969
QY	102	-----	101
Db	12970	KYGVGPVFTPEPAIAKNPYDPPGRCDPPVISNITTKDHTVSWKPPADGGSPITGYLLEK	13029
QY	102	-----	101
Db	13030	RETVANNWTKVNRKPIERTLKATGLOBOTEYEFRTVAINKAGPKPSDASKAAYARDPO	13089
QY	102	-----	101
Db	13090	YPPAPPAPFKVYDTRSSVSLSGKPAYDGGSPIIIGYLVEVKRADSNDWVRNCLPONLOK	13149
QY	102	-----	101
Db	13150	TRPEVTGLMEDTOYQPRVYAVNKGISDPSDVPDKHYPKDILIPPEGEHADLRKTLILR	13209
QY	102	-----	101
Db	13210	AGVTMRLYVPVKGRRPPPKITWSKPNVNLDRIGLDIKSTDFTDFLRCENVNKYDACKYIL	13269
QY	102	-----	101
Db	13270	TLENSCGKKEYTIVVKVLDTPGPPINVTVKESDAYSVTVBPP1IDGGSP1INNVOKR	13329
QY	102	-----	101
Db	13330	DAERKSWSTVTTTECSKTSFRPNLEEGKSYFPRVFAENEYIGDPGETRMAVKASOTPGP	13389
QY	102	-----	101
Db	13390	VVDLKVRVSXSSCSIGWKPHSDGSR1IGYVVDVTEENKQWRVMKSLSLQYSAKDLT	13449
QY	102	-----	101
Db	13450	EGKEYTFRVSAENENGEGTPEITVVARDVAPDLKGLPCLCYLAKENSFRLKIP1	13509
QY	102	-----	101
Db	13510	KGKPAFVSWKKGEDPLATDTRVSVSESSAVNTTLIVYDQCKSAGKYTITLKNVAGTKEG	13569
QY	102	-----	101
Db	13570	TISIKVVGKPGIPTGPIKPFDEVTAEMTLKNAPPKDDGGSEITNYILEKRDVNNKHWTC	13629
QY	102	-----	101
Db	13630	ASAVQKTTFRVTRLHEGMEYTFRVAENKYGVGELKSEPIVARHPDFVPDAPPPNIVD	13689
QY	102	-----	101

Db	13690	VRHDSVSLTWTDPKKTGGSPITGVHLEPKERNSLWKANKTPIRMDFKVTGLTEGLE	13749
Qy	102	-----	101
Db	13750	EFRVMAINLAGVGKPSLPSEPVVALDPIDPPGKPVPLNTRHTVTLLWKAKPEVTG	14889
Qy	102	-----	107
Db	13810	GYIVEKRDLPKSKWMKANHVPECAFTVTDLVEGGKVEFRIRAKNTAGASAPSESTET	14949
Qy	102	-----	107
Db	13870	IICKDEYEAPTIVLDPTIKDGLTIKAGDTIVLNAISILGKLPKSKWSKAGDIRPSDIT	15009
Qy	102	-----	107
Db	13930	QITSTPTSMILTIKYATRKGAGEYITATNPFGTKEVHKVTVLDVPGPPGVSEISVSA	15069
Qy	102	-----	107
Db	13990	EKATLTWTPPLEDGGSPISKYLEKRETSRLLTWTVVSEDIQSCRHVATKLIQGNIEYIPRV	15129
Qy	102	-----	107
Db	14050	SAVNHYGKEPVQSEPVKMDRFGPPGPEKPEVSNVTNTATVSWKRPVDDGGSEITGY	15189
Qy	102	-----	107
Db	14110	HVERREKSLRWRAIKTPVSDLRCKVTGLOQEGSTYEFVRSAENRAGIGPPSEASDSVLM	15249
Qy	102	-----	107
Db	14170	KDAAYPPGPPSNPHVTDITTKSASLAWGKPHYDGGLEITGYVVEHQVDEAWIKDITGT	15309
Qy	102	-----	107
Db	14230	ALRITQFVVPDLQTEKYNFRISAINDAGVGPVDPVEIVEREMAPDFELDAELRRTL	15369
Qy	108	-----	107
Db	14290	VVRAGLSIRIFVPIKGRPAPEVTWKDNLNKNRANIENTESFTLLIPECNRYDTGKFV	15429
Qy	108	-----	107
Db	14350	MTIENPAGKSGFVNVRLDTPQVNLRLRPTDITKDSVTLHWDPLIDGGSRTITNVIVEK	15489
Qy	108	-----	107
Db	14410	REATKRSYSTATTCKCHKCTYKVTGLSEGCEYFFRVMAENEYIGIPEPTETTEPVKASEAPS	15549
Qy	108	-----	107
Db	14470	PPDSLNMIDITKSTVSLAWPKPKHGGSKIYGVIEAQRKGSQDQWTHITTVKGLBCVVRN	15609
Qy	108	-----	107
Db	14530	LTEGEYTFQVMAVNSAGRSAPRESRPVIVKEQTMPELDLRCIYQKLVIAKAGDNIKVE	15669
Qy	108	-----	107
Db	14590	IPVLRPKPTVTKWKGQDILKQTVQVNFETTATSTILNINECVRSDSGPYPLTARNIVGE	15729
Qy	108	-----	107
Db	14650	VGDVITIQVHDIQGPPTGPIKPFDESSDFVTFSDPPENDGGVPISNYVVMQTDSTTW	15789
Qy	108	-----	107
Db	14710	VELATTVIRTYKATRLTTGLEQFRVKAQNRVGVGPGITSAMIYANYPKVPGPPTQ	15849
Qy	108	-----	107
Db	14770	VTAVTKDSMTISWHEPLSDGSGPILGYHVERKERNILQWTVSKALVPGNIFKSSGLTDG	15909
Qy	108	-----	107
Db	14830	IAYEPRVIAENMAGSKSPKSEPLALDPIDPPGKPVPLNTRHTVTLLWKAKPEVTG	15969
Qy	108	-----	107
Db	14890	KITSYIVEKRDLPNGRWLKNFNSILENEFTVSGLTEDAAEFVRVIKAKNAAGISPPSEP	15009
Qy	108	-----	107
Db	14950	SDAITCEDDVEAPKIKVDVKFDVTILKAGEAPRLEADVSGRPPPTMWSKDGKELEGT	15069
Qy	108	-----	107
Db	15010	KLEIKIADFSTNLVNKSDSTRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPPPEGLAVTE	15129
Qy	108	-----	107
Db	15070	VTSEKCVLSWFPPLDDGGAKIDHYIVQKRETSRLANTNVASEVQVTKLVTKLLKGNEYI	15189
Qy	108	-----	107
Db	15130	FRVMAVNYGVGELESEPVLA VNPYGPDPKPNPEVTITKDSMVVCMGHGHPDSDGSEI	15249
Qy	108	-----	107
Db	15190	INVIVERRRKAGORWIKCNKKTLDLRYKVSGLTEGHEVEFRIMAENAGISAPSTSPF	15309
Qy	108	-----	107
Db	15250	YKACDVTFKPGPPGNPRVLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEDEMQIVT	15369
Qy	108	-----	107
Db	15310	PPAGLKATSYITIGLTENQEVKIRIYAMNSEGLGEPALVPGTPKABDRMLPPEIELDADL	15429
Qy	108	-----	107
Db	15370	RKVITIRACCTLRFLVPIKGRPDPEVKWARDHGESLDKASIESASSYTLILVGNVRPDS	15489
Qy	108	-----	107
Db	15430	GKYILTVENSSGSKSAFNVVRVLDTPGPPQDLKVKEVTKTSVLTWDPPLLDGGSKIKNY	15549
Qy	108	-----	107
Db	15490	IVEKRESTRKAYSTVATNCHKTSMKVQLQEGCSYFRVLAENEYIGLPAETAESVKAS	15609
Qy	108	-----	107
Db	15550	ERLPPGKITLMDVTRNSVLSWEKPEHGGSRILGVIVEMQTKGSKWATCATVKVTEA	15669
Qy	108	-----	107
Db	15610	TITGLIOGEYSFRVSAQNEKGISDPQLSVPIAKOLVIPPAPKLLPNTFTVLAGEDLK	15729
Qy	108	-----	107
Db	15670	VDVPEIGRPTPAVTHWKDNVPLKQTRVNAESTENNLSLLTIKACREDVGHVYVVKLTNSA	15789
Qy	108	-----	107
Db	15730	GEAIBTLNVILDKPGPTGPVKMDEVTAISITLSWGPYPKYDGGSSINNYIVEKRDSTTT	15849
Qy	108	-----	107
Db	15790	TWQIVSATVARTTIKACBLKTGCEYQFRIAAENRVGKSTYLSNSEPTVAQYFPKVPGPPTQ	15909
Qy	108	-----	107
Db	15850	PVVTLSSRDSMEVQWNEPISDGGSRVIGYHLERKERNILMWKLNKTPIPQTKFTTGLE	15969
Qy	108	-----	107

QY 108 ----- 107
Db 15910 EGVEYEFVSAENIVGIGKPSKVSECVYARDPCDPGRPEAIIIVTRNSVTLOWKKPTYDG 15969
QY 108 ----- 107
Db 15970 GSKITGYIVEKKELPEGRMKASFTNIIDTHFEVTGLVEDHRYEFVRVARNAAAGVFSEPS 16029
QY 108 ----- 107
Db 16030 ESTGAITARDEVDPRISSMDPKYKDITIVHAGESFKVDADIYGKPIPTIQWIKGDQELSN 16089
QY 108 ----- 107
Db 16090 TARLEIKSTDATSLSVKDAVRDYSNGYILKAKNVAGERSVTYVNVKVLDRPGPEGPVVI 16149
QY 108 ----- 107
Db 16150 SGVTAECTLANPPLQDGGSDIINYIVERRETSRLVWTVVDANVQTLSCVKTKLEGNE 16209
QY 108 ----- 107
Db 16210 YTFRIMAVNKYGVGEPLSEBPVAKNPVVDPAPKAPEVTTVTKDSMIVVWERPASDGGG 16269
QY 108 ----- 107
Db 16270 EILGYVLEKRDKEGIRWTRCHKRLIGELRLRVLTGLIENHDEYEFVSAENAAAGLSEPPSS 16329
QY 108 ----- 107
Db 16330 AYQKADPIYKGPBNPKVIDITRSSVFLSWKPIYDGGCEIOGYIVEKCDVNVGEWTM 16389
QY 108 ----- DL 109
Db 16390 CTPTGINKTNIEVEKLEKHEYNFRICAINKAGVGEHADVPGPPIIVEEKLEAPDIDL 16449
QY 110 EL ----- 111
Db 16450 ELRKIINIRAGGSURLFVPIKGRPTPEVKMGKVDGEIRDAIIDVTSSFTSLVDNRY 16509
QY 112 ----- 111
Db 16510 DSGKYLTLNSSGTSKSAFVTVRVDTPSPVNLKVTEITKDSVITWEPPLDGGSKIK 16569
QY 112 ----- 111
Db 16570 NYIVEKREATRKSAAVVTNCHKNWKIDQLQEGCSYFVRVTAENEYIGLPAQTADPIK 16629
QY 112 ----- 111
Db 16630 VAEVPPPGKITVDDVTRNSVLSWTKPEHDGGSKIIOYIVEMQAKHSEKWCARVKSL 16689
QY 112 ----- 111
Db 16690 QAVITNLTOGEYLFVRVAVNEKGRSDPRSLAVPIVAKDLVIEPDVKPAFSSYSVQVQGD 16749
QY 112 ----- 111
Db 16750 LKMEVPISCRPKPTITWKDGLPKQTTRINVTDSLDITLSIKETHKDDGGQYGITVAN 16809
QY 112 ----- 111
Db 16810 VVGOKTASIEIVTLDKDPKPGKPVKFDVDSAESITLSWNPPLYTGGCOITNYIVQKRD 16869
QY 112 ----- 111
Db 16870 TTWMDVVSATVARTTKVTKLTGTGYQFRIFAENRYGQSFALSDPIVAQYYPKEGPP 16929
QY 112 ----- 111
Db 16930 GTPFATAISKDSMVQIOMHEPVNNGGSPVIGYHLERKERNILWTKVNTIIHDQFKAQN 16989
QY 112 ----- 111

Db 16990 LEEGIEYEFVYAENIVGVGKASKNSSECVYARDPCDPGTPEPIMVKRNEITLOWTKPVY 17049
QY 112 ----- 111
Db 17050 DGGSMITGYIVEKRDLPDGRMKASFTNVIETQFTVSGLTEDORYEFVRIAKNAAGAIK 17109
QY 112 ----- 111
Db 17110 PSDSTGPITAKDEVELPRISMDPKFRDTIVVNAGETFRLEADVHGKPLPTIEMLRGDK 17169
QY 112 ----- PGNR 116
Db 17170 ESARCEIKNTDFKALLIVKDAIRIDGGQYILRASNVAGSKSPVNVKVLDRPGPEGPV 17229
QY 117 QVRGVT ----- OLGG ----- 126
Db 17230 QVTGVTSEKSLTWPPLQDGGSDISHYVVEKRETSRLAWTVASEVVTNSLKVTKLLEG 17289
QY 127 ----- 126
Db 17290 NEYVFRIMAVNKYGVGEPLSEAPVLMKNPFVLPGPCKSLEVTNIAKDSMTVCWNRPSDG 17349
QY 127 ----- 126
Db 17350 GSEIIGYIVEKRDGRGIRWIKCNKRITDLRLRVLTGLTEDHEYEFVSAENAAAGVGEPS 17409
QY 127 ----- ACSP ----- 130
Db 17410 ATVYKACDPVFKPGPPTNAHIVDTTKNSITLWCKPIYDGGSEILGYVVEICKADEEW 17469
QY 131 ----- 130
Db 17470 QIVTQTLGRLVTRFISKLTEHQYKIRVCALNKVGLGEATSVPGTVKPEDKLEAPDL 17529
QY 131 ----- TWS ----- 133
Db 17530 DSELKGIWVAGGSARIHIFPKGRPMPEITWSREEGETDKVQIEKGVNYTQLSIDNCD 17589
QY 134 ----- 133
Db 17590 RNDAGKYLKLENSGSKSAFVTVKVLDTPGPPQNLAVKEVRKDSAFVWEPP1IDGGAK 17649
QY 134 ----- 133
Db 17650 VKNYVIDKRESTRKAYANVSSKCSKTSFKVENLTEGAIYFVRMAENEFVGVPVETVDA 17709
QY 134 ----- 133
Db 17710 VKAAEPPSPGKVTLTVDVSQTSASLWMEKPEHDGGSRLGYVYVEMQPKWKSIVAESK 17769
QY 134 ----- 133
Db 17770 VCNVVTGLSSGOYQFVRKAYNEKGS DPRVLGVPVIAKDLTIQPSLKL PFTYSIOAG 17829
QY 134 ----- 133
Db 17830 EDLKIEIPVIGRPRPNI SWVKDGEPLKQTRVNVETATSTVLHIKEGNKDDFGKYTVTA 17889
QY 134 ----- 133
Db 17890 TNSAGTATENLSVILEKPPPGVPVRFEVSADFFVISWEPYPAYTGGCOISNYIVEKRD 17949
QY 134 ----- 133
Db 17950 TTTTTHMVSATVARTTIKITKLTGTGYQFRIFAENRYGKSAPLDSKAVIVQYPFKEPG 18009
QY 134 ----- 133
Db 18010 PPGTPTVTSISKDMLVQWHEPVNDGGTKIIGYHLEQEKERNILWTKLNTPIQDTKFKT 18069
QY 134 ----- 133

Db	18070	TGLDEGLEVEFKVSAENIVIGICKSVSECFVARDPCDPPGRPEAIVITRNNVTLKWKP	18129
Qy	134	-----	133
Db	18130	AYDGGSKITGYIVEKKDLPGRWKASFVNLETFETVSGLVEDORYEPRVIARNAAGNF	18189
Qy	134	-----	133
Db	18190	SEPSDSSGAITARDEIDAPNASLDPKYKDVIVVHAGETVLEADIRGKPIPDVWMSKDGK	18249
Qy	134	-----	133
Db	18250	ELEETAARMEIKSTIGKTLVVKDCIRTDGOVILKSLNVGCTKSIPITVKVLDPRGSP	18309
Qy	134	-----	133
Db	18310	GPLKVTGVTAEKCYLAWNPLOGGANISHYIIIEKRETSRLSWTQVSTEVALNKKVKL	18369
Qy	134	-----	133
Db	18370	LPGNEVIFRMAVNKYGIGEPLESPVTACNPYKPPGSTPEVSAITKDSMVVTWARPV	18429
Qy	134	-----	133
Db	18430	DDGGTEIEGYILEKRDKEGVRWTKCNKKTLDLRLRVTLGLTECHSYEFRAAENAAGVE	18489
Qy	134	-----	133
Db	18490	PSEPSVFRACDALYPPGPPSNPKVTDTSRSSVSLAWSKPIYDGGAPVKGYVVEVKEAAA	18549
Qy	134	-----	133
Db	18550	DEWTTCTPTGLOGKQFTVTKENTENFRICAINSEGVEPATLPGSVVAQERIEPPE	18609
Qy	134	-----	133
Db	18610	IELDADLRKVVLRASATLRLFVTIKGRPEPEVKEAGELTDRAQIEVTSFTMLVD	18669
Qy	134	-----	133
Db	18670	NVTRFDSGRYNLTLENNSGSKTAFNVVRVLDSPAPNLTIREVKDVSVTLSEPPPLDG	18729
Qy	134	-----	133
Db	18730	GAKITNIVIEKRETRKAYATITNCTKTTFRIENLQEGCSYYFRVLASNEYGIGLPAET	18789
Qy	134	-----	133
Db	18790	TEPVKUSEPPLPRGRVTLVDVTRNTATIKWKPESDGGSKITGYVVMQTKGSEKWSCT	18849
Qy	134	-----	133
Db	18850	QVKTEATISGLTAGEEYVFRVAANKEGRSDPRQLGVPVIARDIEIKPSVELPFHTFNV	18909
Qy	134	-----	133
Db	18910	KAREQLKIDVPFKRPOATVNRKDGQTLKETTRVNVSSKTVTSLSIKEASKEDVGTYE	18969
Qy	134	-----	133
Db	18970	LCVNSAGSITPITIIVLDRPGPPGPIRIDEVSCDSITISWNPPEYDGGCQISNYIVEK	19029
Qy	134	-----	133
Db	19030	KETTSTTHWIVSOAVARTSIKIVRLTTGSEYQFRVCAENRYGKSSYSESSAVVAEYFPSP	19089
Qy	134	-----	133
Db	19090	PGPCTPKVHATKSTMLVTWQVPVNDGGRVIGYHLEYKERSILMSKANKILIADTV	19149
Qy	134	-----	133
Db	19150	KVSGLDEGLMEYRVAENIAGICKSKCEPVPARPCDPGQPEVTNITRKSVSUKWS	19209
Qy	134	-----	133
Qy	134	-----	133
Db	19210	KPHYDGGAKITGYIVERRELPGDGRWLKCNNTNIQETYFEVTELTEDORYEPRVFARNAAD	19269
Qy	134	-----	133
Db	19270	SVSEPSSESTGPIIVKDDVEPRVMDVKPRDVIIVVAKAGEVLKINADIAGRPLPVISWAKD	19329
Qy	134	-----	133
Db	19330	GIEIEEBARTETIISTDNHNTLLTVKDCIRRDITGQVLTLLKNVAGTRSVAVNCKVLDKPGPP	19389
Qy	134	-----	133
Db	19390	AGPLEINGLTAEKCSLSWGRPOEDGADIDYYHRRKKRETSHLAWTICEGELQMTSCVKTK	19449
Qy	134	-----	133
Db	19450	LLKNEYIFRVTGVNKYGVGEPLSVAKALDPFTVPSPPTSLEITSVTKESMTLCHSRP	19509
Qy	134	-----	133
Db	19510	ESDGSSEISGYIIERREKNSLRWVRNKKPVYDLRVKSTGLREGCEYEYRVAENAAAGLS	19569
Qy	134	-----	133
Db	19570	LPSETPLIRAEDPVFLPSPSPKPIVDSGKTTITIAWVKPLPDGGAPITGTYTVEYKKSD	19629
Qy	134	-----	133
Db	19630	DTDMKTSIOSLRGTEYITISGLTTGAEYVFRVKSVNKGASDPSDSDPOIAKEREPEPLF	19689
Qy	134	-----	133
Db	19690	DIDSEMRKTLIVKAGASFTMTVPFRGRVPNVLWSKPDTRLTRAYVDTTDSRTSLTIEN	19749
Qy	134	-----	133
Db	19750	ANRNDSGKYTLTIQNVLSAASLTLLVVKVLDTPGPPTNITVQDVTKESAVLSWDVPENDGG	19809
Qy	134	-----	133
Db	19810	APVKYHIKREASKAWVSVTNNCNRLSYKVTNLOEGAIYYFRVSGENEFVGIPAEK	19869
Qy	134	-----	133
Db	19870	EGVKITEKPSPEKLGVTISIKDSVSLTWLKPHEHDGSRIVHYVVEALEKGQKNWVKCAV	19929
Qy	134	-----	133
Db	19930	AKSTHHVVSGURENSEYFPRVFAENQAGLSDPRELLLPVLKEQLEPPEIDMKNFPSSHV	19989
Qy	134	-----	133
Db	19990	YVRAGSNLKVLDIPISGKPLPKVTLRSRQVPLKATMRPNTETAEENLTINLKESVTADAGR	20049
Qy	134	-----	133
Db	20050	YEITAANSSGTTKAFINIVLDRPGPTGPVVISDITEESVTLKWEPPKYDGGSOVTNYI	20109
Qy	134	-----	133
Db	20110	LLKRETSTAVMTEVSATVARTMMKVMKLTGTGEYQFRIKAENRFGISDHIDSACVTVKLP	20169
Qy	134	-----	133
Db	20170	YTTGPPSTPWNTVNTRESITVGMHEPVNSGSAVVGYHLEMKDRNSILWQKANKLVIRT	20229
Qy	134	-----	133
Db	20230	THPKVTIISAGLIYEFVRVAENAAVGKPSHPSEPLAIDACEPPRNVIRITDISKNSVSL	20289

Qy 134 ----- 133
Db 20290 SMOQAPDGGSKITGIYIVERDLDPGRWTKASFTNVTOFTISGLTONSQYSEFRVARN 20349
Qy 134 ----- 133
Db 20350 AVGSISNPSEVGPITCIDSYGGPVIDLPLEYTEVVKYRAGTSVKLRAGISGRPAFTIEW 20409
Qy 134 ----- 133
Db 20410 YKODKELQTNALVCVENTTDLASILKADRLNSGCYELKLRNAMASASATIRVOILDKP 20469
Qy 134 -----CLIT 137
|:|
Db 20470 GPPGGPIEFKVTVAEKITLLWRPADDGGAKITHYIVKRETSRVVMSVSEHLEECIIT 20529
Qy 138 ----- 137
Db 20530 TTKIIGNEVIFRVAVNKYIGIEPLESDSVVAKNAFVTPGPGIPEVTKITKNSMTVMW 20589
Qy 138 ----- 137
Db 20590 SRPIADGSDISGYFLEKRDKSLGFWKLVKETIRDTRQKVTGLTENSQYRVCAVNA 20649
Qy 138 ----- 137
Db 20650 QGPFSEPFYKAADPIDPPGPAKIRIADSTKSSITLWGSKPVDGGSVAVTGYVVEIR 20709
Qy 138 ----- 137
Db 20710 QGBEEWTVSTKGEVTRTEYVVSNLKPGVNYFRVSNVNCAGOGPIEMNEPVQAKDIL 20769
Qy 138 ----- 137
Db 20770 EAPEIDLVALTSIAKAGEDVQVLIPFKGRPPPTVTRKDEKNLGDARSIENTDSS 20829
Qy 138 ----- 137
Db 20830 SLLTIPOVTRNDTGKYLITIENGVEPKSVTSVKVLDTPAACQKLQKHVSRGTVTLW 20889
Qy 138 ----- 137
Db 20890 DPPLIDGSPINIYVIEKRDATKRTSVVSHKCSSTSKLIDLSEKTPFFFRVLAENEIG 20949
Qy 138 ----- 137
Db 20950 ICEPCETEPVKAEEVPAPIRDLMSKDKSTSVILSWTKPDPDGGSVITEYVVERKKG 21009
Qy 138 ----- 137
Db 21010 QTMSHAGISKTCEIEVSQLEKQSVLEFRVFAKNEKGLSDPVTIGPITVKELIITPEVDLS 21069
Qy 138 ----- 137
Db 21070 DIPGAQVTVRIGHNVHLELPYKGPKPSISMLKDGLPLKESEFVRFSKTENKITLSIKNA 21129
Qy 138 ----- 137
Db 21130 KKEHGKVTVIDNAVCRVAPITVITLGPSPKPKPIRFDKADSVILSDVPEDNGG 21189
Qy 138 ----- 137
Db 21190 GEITCYSIEKRETSQTNWKCSSVARTTFKVPNLVKDAEYQFVRAENRYGVSOPLVSS 21249
Qy 138 ----- 137
Db 21250 IIVAKHQFRIIPGPPGKPIYINVTSDGMSLTWDAPVVDGGSEVTGPHVEKKERNILMOKV 21309
Qy 138 ----- 137
Db 21310 NTSPIGREYRATGLVEGLDYQFVVAENSAGLSPEKFTLAVSPVDPGTPDYIDV 21369
Qy 138 ----- 137

Db 21370 TRETTILKWNPLRDGSGKIVGYSIEKROGNERWRCNFTDVSECOVTVTGLSPCDRVEF 21429
Qy 138 ----- 137
Db 21430 RIIARNAVGTISPPSQSSGIIMTRDENVPPIVEFGEYFDGLIIKSGESLRIKALVOORP 21489
Qy 138 ----- 137
Db 21490 VPRVTWFKDGVIEIKRMWMEITNVLGSTSLFVRDATDRHRCVYVVEAKVSGSAKAEIKV 21549
Qy 138 ----- 137
Db 21550 KVQDTPGVGVGPIRFTNITGEKMTLWMDAPLNDGCAPITHYIIEKRETSRLAWALIEDKC 21609
Qy 138 ----- 137
Db 21610 EAQSYTAIKLINGNEYQFRVSAVNFVGRPLDSDPVVAQIOYTVDPDAPGIPESNITGN 21669
Qy 138 ----- 137
Db 21670 SITLTWARPESDGGSEIOQYILEREKKSSTRWVKVISKRPISETRKFVKTGLTEGNEYBPH 21729
Qy 138 ----- 137
Db 21730 VMAENAGVGPASGISRLIKCREPVNPPGPTVVKVTDTSKTTVSLEWSKPVFDGMEII 21789
Qy 138 ----- 137
Db 21790 GYIIECKTDLGDMHKVNAEACVTRYTVTDLOAGEEYKFRVSAINGAGKSDCEVTGTI 21849
Qy 138 ----- 137
Db 21850 KAVDRLTAPELDIDANFKQTHVVVRAGASIRLFIAVQGRPTPTAVMSKPDNSLSRADIHT 21909
Qy 138 ----- 137
Db 21910 TDSFSTLTVCNCRNDAGKYTLTVENNSGSKSITFTVKVLDTPGPPGPTFKDVTGTSAT 21969
Qy 138 ----- 137
Db 21970 LMDAPLDGGARIHHYVVEKREASRRSHQVISEKCTROI FKVNDLAEGVPPYFRVSAVN 22029
Qy 138 ----- 137
Db 22030 EYGVGEPEMPEIVATEQAPRRRLDVVDTSSSAVLAWLKPDDHSGSRITGYLLEMRO 22089
Qy 138 ----- 137
Db 22090 KGSDLWEAGHTKQLTFTTVERLVEKTEYEFVRKAKNDAGYSEPREAFSSWIKPEOIEPT 22149
Qy 138 ----- 137
Db 22150 ADLTGNTQLITCKAGSPFTIDVPISGRPAPKVTKLEMLRKETDRVSIITTKDRTTLT 22209
Qy 138 ----- 137
Db 22210 VKDSMRGDSGRYFLTLENTAGVKVTSVTVVIGRPGVGTGPIEVSSVSAESCVLWSGEPK 22269
Qy 138 ----- 137
Db 22270 DGGTEITNYIVEKRESGTTAMQLVNSSVKRTOIKVTHLTKYMEYFRVSSSENRFVSKP 22329
Qy 138 ----- 137
Db 22330 LESAPIIAEHFPVPPSAPTPEVYHVHSANAMSIRWEPYHDGSGKIIGYHVEKKERTIL 22389
Qy 138 ----- 137
Db 22390 WKENKVPCLCNKVKVTGLVEGLEQFRTYALNAAGVSKASEASRPTMAQNPVDAPGRPE 22449
Qy 138 ----- 137

Db	22450	VTDVTRSTVSLWSAPAYDGGSKVVGYYIERKPVSEVGDGRWLKCNYYTIVSDNFFVTAL	22509
Qy	138	-----	137
Db	22510	SEGDTYFRVLAKNAAGVISKSESTGPVTCRDEYAPPAELDARLHGLDVTIRAGSDLV	22569
Qy	138	-----	137
Db	22570	LDAVGGKPEPKIHWTKGKELDLCVKVSLQYTGKATAVIKFCDRSDSGKTYLTVKNAS	22629
Qy	138	-----	137
Db	22630	GTKAVSMVMKLDSPGCGKLTSVRTOBKCTLAWSLPQEDGAEITHYIVERRETSRLN	22689
Qy	138	-----	137
Db	22690	WVIVEGECTLSYVVTRLKKNEXYIFRVRANKYGPVPESEPIVARNSTFIPSPGP	22749
Qy	138	-----	137
Db	22750	EEVGTGKEHIIQWTKPESDGGNEISNLYVDKREKESLRWTRVNDYVYDTRLKVTSLM	22809
Qy	138	-----	137
Db	22810	EGCDYQFRVAVNAAGNSEPSESNFISCREPSYTPGPSAPRVVDTTKHSISLAWTKPM	22869
Qy	138	-----	137
Db	22870	YDGGTDIVGVLEMOEKDOWYRVHTNATIRNTEFTVPLKMGOKYSFRVAANVKNKMS	22929
Qy	138	-----	137
Db	22930	EYSESAIEPVERIEIPLELADDLKKTITRAGASLRMLVSVGRPPPVITWSKQGD	22989
Qy	138	-----	137
Db	22990	LASRAIIDTESYLLIVDKVNRDAGKVTIEAENOSGKSATVLVKVYDTPGPCPSVKV	23049
Qy	138	-----	137
Db	23050	KEVSRDSVTITWEIPTDGGAPINNYIVEKREAAAFKTVTTKCSKTLRISGLVEGTM	23109
Qy	138	-----	137
Db	23110	HYPRVLPENIYGICEPCESTDVLSVEPLVPKLEVDVTKSTVLAWEKPLYDGGSL	23169
Qy	138	-----	137
Db	23170	TGYVLEACKAGTERMKVTLKPTVLEHTVTSNEDGEQYLFIRIAQNEKGVSEPRETVTA	23229
Qy	138	-----	137
Db	23230	VTQDLRVLPDIDLSMTPOKTHVPAGRPVELVPIAGRPPPAASWFFAGSKLRESERV	23289
Qy	138	-----	141
Db	23290	VEHTKVAKLTIRETTIRDTGEYLELKNVTGTTSETIKVILDKGPPTGPIKIDEIDA	23349
Qy	142	-----	141
Db	23350	TSITISMEPPELDGGAPLSGVVVEQDAHRPGMLPVSESVTRSTFKFTRLTEGNEYVFRV	23409
Qy	142	-----	141
Db	23410	AATNRFGISYLOSEVIECRSSIRIPGPPELQIFDVSRDGMILTWPYPEDDGGSQVTGY	23469
Qy	142	-----	141
Db	23470	IVERKEVRADRWVRVNVPTMTTRYSTGLTEGLEYEHRVTAINARGSGKPSRPSKPIVA	23529
Qy	142	-----	141
Db	23530	MDPIAPPGKQNPVTDTRTSVSLAWSVPEDEGGSKVTYGILBQMKVQDHEWTKCNTTP	23589
Qy	142	-----	141
Qy	142	-----	141
Db	23590	TKIREYTLTHLPOGAEYRFRVLACNAGGCEPAEVPGTVKVTEMLVEYPOVELDERYOEGI	23649
Qy	142	-----	141
Db	23650	FVRQGGVIRLTIPKIGKPPFICKWTKEGQDISKRAMIATSETHTELVIKEADRGDSGYD	23709
Qy	142	-----	141
Db	23710	LVLENCKGKAAVYKVRVIGSPNSPEGPLEYDDIQVRSVRVSMRPADGGADILGYILE	23769
Qy	142	-----	141
Db	23770	RREVPKAAWYITIDSRVGTSLVVVKGLKENVEYHFRVSAENOFGISPKLKEEPTVTKTPL	23829
Qy	142	-----	141
Db	23830	NPPEPPSNPEVLDVTKSVLSWSRPPKDDGSRVTGYIETKSTDKVVRHNKTOITT	23889
Qy	142	-----	141
Db	23890	TMYTGTGLVPDAEQFRIIAQNDVGLSETSPASEPVVCKDPFKPSQSGELEILSISKDS	23949
Qy	142	-----	141
Db	23950	VTLOWEKEPCDGGKEILGYWVEYRQSGDSAWKSNKERIKDKOFTIGGLLEATEYEPRVF	24009
Qy	142	-----	141
Db	24010	AENETGLSRPRRTAMSIKTKLTSGEAPGIRKEMKDVTTLGEEAQLSCQIVGRPLPIKW	24069
Qy	142	-----	141
Db	24070	YRFKELIQSRKYKSSDGRTHLTVMTEDEDEGVVTCIATNEVGEVETSKLLLOATP	24129
Qy	142	-----	141
Db	24130	QHPGYPLKXYGAVGSTLRHLVHYIGRPVPMTWPHGKQLLONSENITIENTEHTYHL	24189
Qy	142	-----	141
Db	24190	VKNVORVTHAGYKVKVLSNVFGTVDAILDVEIODKPKPTGPIVIEALLKNSAVISWKP	24249
Qy	142	-----	141
Db	24250	PADGGSMITNYVVEKCEKEGAEWQLVSSAISVTTCRIVNLTENAGYVFRVSAQNTFGI	24309
Qy	142	-----	141
Db	24310	SDPLEVSSVWIIKSPFEKPGAPGKPTITAVTKDSCVVAWKPASDGGAKIRNYYLEKREK	24369
Qy	142	-----	141
Db	24370	KONKWI SVTTEIRETVFSVKNLIEGLEVEFRVKCNELGSESEISEPIPKSDVPIQ	24429
Qy	142	-----	141
Db	24430	APHFKEELNLRVQSNATLVCKVTHPKPIVKWYRQGEIADGLKYRIOEFKGGYHQ	24489
Qy	142	-----	141
Db	24490	LIASVTDDATVYQVRATNOCSSVSGTSASLEVEVPKIHLPKLTLEGMVHALRGEVVS	24549
Qy	142	-----	141
Db	24550	IKIPSGKPDPIVTHQKQDLDNNHGYQVIVTRSTSLVPNGVVERKODAGEVYVCAKNR	24609
Qy	142	-----	141
Db	24610	FGIDQKTVELDVADVPDPPRGVKVSDASRDSVNLTWTEPASDGGSKITNYIVEKCATTAE	24669

Db 730 HIKTDDQGMHISSQIKKTTDLTTERLVHVDKRPRTASPHFTVSKISVPKTEHCYEASIA 789
Qy 19 ----- 18
Db 790 GSAIATLOKELSATSSAQKITKSVKAPTVPKSETRVRAEPTPLPQFPFADPTDYKSEAG 849
Qy 19 ----- 18
Db 850 VEVKKEGVSGITGTVREERFEVLHGREAKVTETARVPVEIIVTPTPLVSGLKNTVI 909
Qy 19 -----FQ-----AHQEDTERYV- 30
Db 910 EGESVTLCHISGYPSPVTHWYREDYQIESSIDQITFQSGIARLMIREAFEDSGRFTC 969
Qy 31 ----- 30
Db 970 SAVNEAGTVSTCYLAVQVSEFEKETTAVENTEKTTEERKREFESRDVVMTDTSLTEBQAG 1029
Qy 31 ----- 30
Db 1030 PGEPAAPYFITKPVVQKLVGGSVVFGQVGNPKPHVYWKSGVPLTTGYRYKVSYNKQ 1089
Qy 31 ----- 30
Db 1090 TGECKLVISMTFADDAGEYTIIVRNKHGETSASASLLEADYELLMKSQEWMLYQTVTA 1149
Qy 31 ----- 30
Db 1150 FVQEPKVGETAPGFVYSEYEKEQALIRKMAKDTVVVRYVEDQEFHISSEERLI 1209
Qy 31 ----- 30
Db 1210 KEIERYIINTLEELLEDGEBKMAVDISEAVESGFDLRINKYRIILEGMCVTFCHKMS 1269
Qy 31 ----- 30
Db 1270 GYPLPKIANYKDGRIKHGERYQMDFLQDGRASLRIPVVLPEDEGIYAFASNIKGNAC 1329
Qy 31 ----- 30
Db 1330 SGKLYVEPAAPLGAPTYIPTLEPVSIRLSRSPRSVSRPSPARMSPARMSPA 1389
Qy 31 ----- 30
Db 1390 RMSPGRRLEETDESQERLYKPVFLKPVSKFCLEGQATARFDLKVVGRPMPETFWHDQO 1449
Qy 31 ----- 30
Db 1450 QIVNDYTHKVVIKEDGTQSLIIVPATPSDGEWTVVAQNAGRSSISVILTVEAVEHQVK 1509
Qy 31 -----LTNLI----- 36
Db 1510 PMFEVKLNKVNKEGSOLEMKVRATGNPNPDIVWLKNSDIIVPHKYPKIRIEGTKEAAL 1569
Qy 37 ----- 36
Db 1570 KIDTVSODSAWYATATAINKAGROTTTRCKVNVEVEFAEPEPERKLIIPRGTYRAKEIAAP 1629
Qy 37 ----- 36
Db 1630 ELEPLHLRYQEQWEGDLYDKEKQKPPFKKXLTSLRLKRFQPAHFECRLTPIGDPWV 1689
Qy 37 ----- 36
Db 1690 VEWLHDKPLEAANRLRMINEFGYCSLDYGVAYSRSGLITCRATNKYGTDHTSATLIYK 1749
Qy 37 -----GA----- 38
Db 1750 DEKSLVEESQPEGRKGLQRIEELERMAHEGALTGVTTDQKEKQKPDIVLYPEPVRVLEG 1809
Qy 39 ----- 38
Db 1810 ETARFRCRVTGPOPQKVNWYLNQGLIRKSRFRVRVYDGIHLYDIVDCKSYDTGEVKVTAE 1869

Qy 39 ----- 38
Db 1870 NPEGVTEHKVKLEIQOEDFRSVLRRAPEPRPEFHVHBPGLQFEVQKVDRPVDTTETKE 1929
Qy 39 -----ELLR--- 42
Db 1930 VVKLKAERITHEKVPSESEELRSKPKRRTEEGYEAITAIVELKSRKKDSYEELURKTK 1989
Qy 43 ----- 42
Db 1990 DELLHWKTLETEEEKALABEGKITIPTPKPKIELSPSMEAPKIFERIOSQTVOGSDA 2049
Qy 43 -----DPS----- 45
Db 2050 HFRVRVVGKPDPECEWYKNGVKIERSDRIYWYWPEDNVCELVIDVTAEDSASIMVKAIN 2109
Qy 46 ----- 45
Db 2110 IAGETSSHAFLVQAKOLITFTOELQDVVAKEDTMTATPECETSEPFVKVKYKDGMEVH 2169
Qy 46 ----- 45
Db 2170 EGDYRMHSDRKVHFLSILITDTSDAEDYSCVLVEDENVKTTAKLIVEGAVVEFVKELQD 2229
Qy 46 ----- 45
Db 2230 IEVPESYSGELECIVSPENIEGKWHYHNDVELKNGKYTITRRGRQNLTVKDVTKEDQGE 2289
Qy 46 ----- 45
Db 2290 YSFVIDGKTTCKLKMKPRPIAIIQGLSDQKCEGDIVQEVKVSLESVEGVMMKQGOEV 2349
Qy 46 ----- 45
Db 2350 QPSDRVHIVIDKQSHMLLIEDMTKEDAGNYSFTIPALGLSTSGRVSVYSDVITPLKDVN 2409
Qy 46 ----- 45
Db 2410 VIEGTKAVLECKVSPDVTSVKMYLNDEQIKPDDRVOAIVKGTQKQRLVINRTHASDEGYP 2469
Qy 46 ----- 45
Db 2470 KLIQVRETNCNLSEVEKIKIIRGLRDLTCTETQNVVFEVLSHSGIDVLNMFNFKDEIKPS 2529
Qy 46 ----- 45
Db 2530 SKYKIEAHGKIYKLTVLNMMKDDGKVTYFAGENITSGKLTVAGGAISKPLTDQTVAESQ 2589
Qy 46 ----- 45
Db 2590 EAVFECEVANPDSKGEWLRDQKHLPLTNNIRSESDGHKRRLIITAATKLDDDIGEYTYKVAT 2649
Qy 46 ----- 45
Db 2650 SKTSAKLKVEAVKIKKTLKNLTVTOTDAVFTVELTHPNVKGQVQWIKNGVGVLESNEKYAI 2709
Qy 46 -----LGAQFRVHL---VQWV----- 58
Db 2710 SVKGTIYSLRIKNCIAIVDESIVYFRLGRLGASARLHVETVKIHKPKPDVTALENATVAFE 2769
Qy 59 ----- 58
Db 2770 VSVSHDTPVVKWFHKNVEIKPSDKHRLVSEKRVKHLMLQNLSPSDAGEYTAIVGOLECKA 2829
Qy 59 ----- 58
Db 2830 KLFVETLHITKMKNIIEVPETKTASFECEVSHFNVPNSMWLKNQVEIEMSEKPKIIVVOGKL 2889
Qy 59 ----- 58
Db 2890 HQLIIMNTSTEDSAEYTFVCGNDQVSATLTVTPIMTSMLKDINAEEKDITTFEVTVNYE 2949

QY 59 ----- 58
Db 2950 GISYKWLKNGVEIKSTDKCOMRTKLLTHSLINIRNVHFGDAADYTFVAGKATSTATLYVEA 3009
QY 59 ----- 58
Db 3010 RHIEFRKHIDIKVLEKRAMFEVSEPDITVQMKDDQELQITDRIKIOKEYVHRL 3069
QY 59 ----- 58
Db 3070 IPSTRMSDAGKYTVVAGGNVSTAKLFVEGRDVRIRSIKKEVQVIEKORAVVEFVNEDDV 3129
QY 59 ----- 58
Db 3130 DAHWYKDGIEINFQOERHKYVVERIRHMFISETRSDAGEYTFVAGNRSSVTLYVNA 3189
QY 59 ----- 58
Db 3190 PEPPQVLQELQVTVOSGKPARFCAVISGRPOPKISWYKEEQLLSTGFKCKFLHDCQEYT 3249
QY 59 ----- 58
Db 3250 LLLIEAPPDAANYTCEAKNDYGVAATTSASLSVEPEVSPDQEMPVYPALITPLQDTV 3309
QY 59 ----- 58
Db 3310 TSEGOPARFCVRSGTDLKVSWYKDKKIKPSRFRMTQFEDTYQLEIAEAYPEDEGTYT 3369
QY 59 ----- 58
Db 3370 FVASNAVQVSSANLSLEAPESILHERIEQEIEMEMKEFSSSFLSABEEGLHSAELQLS 3429
QY 59 ----- 58
Db 3430 KINETLELLSESPVYSTKFDSEKEGTGPIFIKEVSNADISMGDVATLSVTVIGIPKIQ 3489
QY 59 ----- 58
Db 3490 WFFNGVLLTPSADYKVFVFGDDHSLIILFTKLEDEGEYTCNASNDYGTICSAYLKINSK 3549
QY 59 ----- 58
Db 3550 GEGHKDTETESAVAKSLEKLGPGCPHFLKELKPIRCAOGLPAIFEYTVVGEPAPTVWF 3609
QY 59 ----- 58
Db 3610 KENQOLCTSVYTTIHNPNNGSGTFIVNDPQREDSGLYICKAENMLGESTCAAELLVLED 3669
QY 59 ----- 58
Db 3670 TDMTDTCKAKSTPEAPEDFPQTPKGPVAVBALDSEQEIATFVKDITLKAALITEENQOL 3729
QY 59 ----- 58
Db 3730 SYEHIKANELSSQLPLQAQELQILEODKLTPESTREFLCINGSIHFPKLEPSNQL 3789
QY 59 ----- ILTEPEG- 65
: |||
Db 3790 QIVSQKTFSGEGLMPEEPETOAVLSDTEKIFPSAMSIEQINSITVEPLKTLAEPEGN 3849
QY 66 ----- 65
Db 3850 YPQSSIEPPMHSYLTSVAEVLSPKEKTVSDTNREQRTLOKQEAQASALILSOLAECHV 3909
QY 66 ----- 65
Db 3910 ESLQSPDVMISQVNYEPLVPSEHSCTEGGKILIESANPLENAGQDSAVRIEKGSLRPPL 3969
QY 66 ----- 65
Db 3970 ALEEKVLLKEHSDNVMPDQIIIESKREPVAIKKVOEVOGRDILLSKESLLSGIPESQR 4029
QY 66 ----- 65

Db 4030 LNLKIQICRALQAQAVASEQPOLFSEWLRNIEKVEAVNITQEPRHIMCMYLVTSKSVT 4089
QY 66 ----- 65
Db 4090 EEVTHIIIEDVDPQMANLKMELDALCAIYBEEIDILTAEGPRIQOGAKTSLQEEMDSFG 4149
QY 66 ----- 65
Db 4150 SQKVEPITEPEVESKYLISPEEVSFVQSRVKYLDATPTKGVASAVVSDKQDES LKP 4209
QY 66 ----- 65
Db 4210 SEEKESSSESGTEBVAIVTKIQEAGGFIKEDGPMIHTPLVDTVSEEGDI VHLTTSITNA 4269
QY 66 ----- 65
Db 4270 KEVNMWFENKLVPSDEKFKCLQDQNTYTLVIDKVNTHDHOGEYVCEALNDSGKTATS AKL 4329
QY 66 ----- -APNI- 69
: |||
Db 4330 TVVKRAAPVIRKRKIEPLEVALCHLAKFTCEIOSAPNVRFQWFKAGREIYESDKCSIRSSK 4389
QY 70 ----- 69
Db 4390 YISSLEILRTQVDCGEYTCASNEYSGSVCTATLTVTVPGEKKVKRLPERKPEPKEE 4449
QY 70 ----- 69
Db 4450 VVLKSVLRRRPEEBEPKVEPKKLEKVKPAVPEPPPKPVEVEVPTVTKRERKIPEPTK 4509
QY 70 ----- 69
Db 4510 VPEIKPAIPLPAPEPKPKPEAEVKTIKPPVPEPEPTIAAPVTVPVVGKAKAKAPKEBA 4569
QY 70 ----- 69
Db 4570 AKPKGPIKGVPKKTPSPTEAERRKLRPGSGGKPPDEAPFTYQLKAVPLKFVKEIKDIL 4629
QY 70 ----- 69
Db 4630 TESEFVGSSAIFELVSPSTAITTWMKOGSNIRESPKHFIAADGDKRKHIIIDVQLSDAG 4689
QY 70 ----- 69
Db 4690 EYTCVURLGNKEKSTAKLVVEELPVRFVKTLSEEVTVVGQPLYLSCELNKERDVVWRK 4749
QY 70 ----- 69
Db 4750 DGKIVVEKGRIVPGVIGLMRALTINDADDTDAGTVTVTVENANNLECSSVVKVVEVIRD 4809
QY 70 ----- 69
Db 4810 WLVKPIRDQHVKPGKTAIFACDIADKTPNIKWFPGYDEIPABPNDKTEILRDGNHLYLKI 4869
QY 70 ----- 69
Db 4870 KNAMPEDIAEYAVEIESGRYPAKLTUGEREVELLKPIEDVTIYKESASFDAEISADIP 4929
QY 70 ----- 69
Db 4930 GOWKLXGELLRPSPTCEIKAEGKRFLLTKVKLDOAGEVLVQALNAITTAILTIVKEIEL 4989
QY 70 ----- 69
Db 4990 DFAVPLKDVTVPERROARFECVLTREANVWSKGPDI IKSSDKFDIIADGKKHILVINDS 5049
QY 70 ----- 69
Db 5050 QFDDGCVYTAEVGKKT SARLFVTGIRLKFMSPLEDQTVKEGETATFVCELSHEKMHVW 5109
QY 70 ----- 69

Db	5110	FKNDAKLHTRVTLISSEKTHKLEMKVTLDDISQIAQVKELSTAQLKYLEADPYFT	5169
Qy	70	-----	69
Db	5170	VKLHKTAVEKDEITLCKVSKDVVKWFKDGEEIVPSPKYSIKADGLRRILIKKADIK	5229
Qy	70	-----	69
Db	5230	DKGEYVDCGTDKTKANVTVEARLIKVEKPLYGVEVFGETAHFEIELESEPDVHCQWKLK	5289
Qy	70	-----	69
Db	5290	GQPLTASPDCEIIEGCKHLILHNCQLGTMGEVSFOAANAKSAANLKVKELPLIFITPL	5349
Qy	70	-----	69
Db	5350	SDVKVEKDEAKPECEVSREPKTFRWLKGTQETGDDRFELIKDGTKHSMVIKSAAFEDE	5409
Qy	70	-----	69
Db	5410	AKYMFADKHTSGKLIIEGIRLKFUTPLKDVTAKEKESAVFTVELSHDNIRVKWFKNDQ	5469
Qy	70	-----	69
Db	5470	RLHTRSVSQDEGKTHSITFKDLSIDDTSQIRVEAMGHSSEAKLTVLEGDPYFTGKLQD	5529
Qy	70	-----	69
Db	5530	YTGVKEDEVILCEISKADAPVKWFKDGEIKPSKNAVIKADGKRMLILKKALKSDIQO	5589
Qy	70	-----	69
Db	5590	YTCDCGTDKTSGLDIEDREIKLVRPLHSVEVMEETARFETISEDDIHANWKLGEAL	5649
Qy	70	-----	69
Db	5650	LQTPDCEIKEEGKIHLVHLNCRLDQTGGVDFQAAVNVKSAHLRVKPRVIGLRPLKDV	5709
Qy	70	-----	69
Db	5710	VTAGETATFDCELSYEDIPVEWYLGKGLPEPSDKVVRSEGKVTTLTRDVKLEDA	5769
Qy	70	-----	69
Db	5770	LTAKDFKTHANLKVKEPPVEFTKPLEDOTVEEGATAVLECEVSRENKVKWFKNGTEILK	5829
Qy	70	-----	69
Db	5830	SKKYEIVADGRVKRLVIHDCTPEDIKTYTCDAKDFKTCNLNVPPHVEFLRPLTDQVR	5889
Qy	70	-----	69
Db	5890	EKEMARFECELSRENKVKWFKDGAETKGGKYDIISKGAVRILVINKCLLDDEAEYSCE	5949
Qy	70	-----	69
Db	5950	VRTARTSGLTVLEEEAVFTKNLANIEVSETDTIKLVCEVSKPGAEVIWYKGDDEIETG	6009
Qy	70	-----	69
Db	6010	RYEILTEGRKILVIONAHLEDAGNVCNRLPSSRTDGVKVKVHELAAEFISKPNLEILEG	6069
Qy	70	-----	69
Db	6070	EKAEFVCSISKESFPQWKRDDKTLES GDKYDVIADGKKRVLVVKDATLQDMGYVVMVG	6129
Qy	70	-----	69
Db	6130	AARAAHLTVIEKLRIVVPLKDRVKEQEVWFNCEVNTREGAKWFRNEAIFDSSKYI	6189
Qy	70	-----	69
Db	6190	ILOKDLVYTLRIRDAHLDDQANVNVSLTNHNGENVKSAANLIVEEEDLRIVEPLKDIETM	6249
Qy	70	-----	69
Db	6250	EKKSVTWCKVKNRLNVLTKWTQNGEEVFPDNRVSRVDKYKHMLTIKDCGPPDSEGEIVT	6309
Qy	70	-----	69
Db	6310	AGQDKSVAELLIIIEAPTEFVEHLEDQTVTEPDDAVFSCLSREKANVKWYRNGREIKEGK	6369
Qy	70	-----	69
Db	6370	KYFEKOGSIHRLIIKOCRLDDECEYACGVEDRKSRARLFEVEIPVEIIRPPQDILEAPG	6429
Qy	70	-----	69
Db	6430	ADVPLAELNKDKVEVQWLRNNMVVQGDQKHQMSEKGIHRLQICDIKPRDQGEYRFAK	6489
Qy	70	-----	69
Db	6490	DKEARAKLEAAPKIKITADQDLVVDVGKPLTMVVPYDAYPKAEAEWFKENEPLSTKTID	6549
Qy	70	-----	69
Db	6550	TTASQTSFRILEAKKDGKGRYKIVLQNHGKAEGFINLKVIDVPGVPRNLEVTEFTFDEGV	6609
Qy	70	-----	69
Db	6610	SLAWEEPLTDGSKIIGYVVERRDIKRKTWVLTADRAESCEFTVTGLOKGVVFLRVSA	6669
Qy	70	-----	69
Db	6670	RNRVGTGPEVETDNPVEARSKYDVPPLNVITITDVNRFGVSLTWEPPEDGGAEITNVV	6729
Qy	70	-----	69
Db	6730	IELRDKTSIRWDNTAMTVRAEDLSATVTVVEGQEVSRVRAQNRIGVCKPSAATPFVKA	6789
Qy	70	-----	76
Db	6790	DIERPSPVNLTSDDQTSQSVOLKWEPLKDGGSPIILGYIIERCEGKDNMIRCNMKLV	6849
Qy	77	-----	76
Db	6850	PELTYKVTGLEKGNKYLYRVAENKAGVSDPSEILGPLTADDAFVEPTMDLSAFKDGLEV	6909
Qy	77	-----	76
Db	6910	IVPNPITILVPSTGYPRPTATWCFGDKVLETCDRVVKMKTLSAYAEVLVSPSERSDKGIY	6969
Qy	77	-----	76
Db	6970	LKLENRVKTSIGEIDVNVVIARPSAPKELKFGDITKDSVHLTWEPDDDDGGSPLTGYVVEK	7029
Qy	77	-----	81
Db	7030	REVSRTWKVMDFTVDLEFTVVDLVQGEYLFVKVARNKCGGEPAYVDEPNVMTSPAT	7089
Qy	82	-----	81
Db	7090	VDPDPNVKWRDRTANSIFLTWDPKNDGGSRIKGYIVERCPRGSKWVACGEPVAETKM	7149
Qy	82	-----	81
Db	7150	EVTGLEEGKWAYRVKALNRQAGSKPSRPTBEOAVDTQEAPEIFLDVKLLAGLTVKACT	7209
Qy	82	-----	81
Db	7210	KIELPATVTKPEPKITWTAKMILKQDKRITIEVNPVKSTVTIIVDSKRSDDGTGYILEAV	7269
Qy	82	-----	81
Db	7270	NVCGRATAVVEYNVLDKPGPPAAFDITDVTNESCULLTWNPPRDDGGSKITVNVVERRATD	7329
Qy	82	-----	81

QY 82 ----- 81
Db 7330 SEVHKLSSVKTDFKATKLIIPKEYIFRVAENMYGVEPVQASPIITAKYQFDPGPP 7389
QY 82 ----- 81
Db 7390 TRLEPSDITKDAVLTWCEPDDGGSPITGYWVERLDPTDKWRCNKMVPKDTTIRVKG 7449
QY 82 ----- 81
Db 7450 LTNKKYFRVLAENLAGPKSKSTEPILIKOPIDPPWPKPTVKDVGKTSVRLNWK 7509
QY 82 ----- 81
Db 7510 PEHGGAKIESYVIEMLTGTDEWVRVAEGVPTTQHLLPGLMEGQESFRVRAVNKAGES 7569
QY 82 ----- 81
Db 7570 EPSESPDLCREKLYPPSPRMLEVINITKNATDLKWTVPKDGSGSPITNYIVEKRDVR 7629
QY 82 ----- 81
Db 7630 RKGHOTVDTTKCTVTPLTEGSLYFRVAENAIQSDYTEIEDSVLAKDTFTTGP 7689
QY 82 ----- 81
Db 7690 PYALAVVDVTKRHVDLKWEPKNDGGRPIORYVIEKKERLGRWVKAGTAGPDCNFRVT 7749
QY 82 ----- 81
Db 7750 DVIEGTEVOFVRAENAGVHPSEPTIELSIDPTSPSPPLDLHVTDAGRKHIAIAWK 7809
QY 82 ----- 81
Db 7810 PPEKNGSPILGYHVCMPCVTEKMRVNSRPIKDLKFVEGVVPDKEYVLRVAVNAI 7869
QY 82 ----- 81
Db 7870 GVSEPSEISENVAKDPCKPTIDLETHDIIIEGKLSIPVPFRAVPVPTVSWHKOGKE 7929
QY 82 ----- 81
Db 7930 VKASRLTMKNDHSAHLEVPKSVRADAGIYITILENKLGSATASINVKVIGLPGPCKDI 7989
QY 82 ----- 81
Db 7990 KASDITKSSCKLTWEPPEFDGTPILHYVLERREAGRRTYIPVMSGENKLSWTVKDLIPN 8049
QY 82 ----- 81
Db 8050 GEYFFRVAKNVKGGEYIELKNPVIAQDPKQPPDPPVDVEVHNPTAEAMTITWKPPLYD 8109
QY 82 ----- 81
Db 8110 GGSKIMGYIIIEKIAKEERWKCNEHLVPILTYTAKGLEEGKEYQFVRAENAGISEPS 8169
QY 82 ----- 81
Db 8170 RATPPTKAVDPIDAPKVILRTSLEVRGDEIALDASISGSPYPTITWIKDENVIIVPEEIK 8229
QY 82 ----- 81
Db 8230 KRAAPLVRRRKGEVQEBEPFVLPLTORLSIDNSKKGESQLRVRDSLDPDHGLYMIKVEND 8289
QY 82 ----- 81
Db 8290 HGIAPACTVSLDTPGPINFVFEDIRKTSVLCKWEPLDDGGSEIINYTLEKKDKTKP 8349
QY 82 ----- 81
Db 8350 DSEWIVVTSLRHCKYSVTKLIEGKEYLFRVAENRFGPPPCVSKPLVAKDPFGPPDAP 8409
QY 82 ----- 81

Db 8410 DKPIVEDTNSMLVKWNEPKDNGSPILGYWLEKREYNSTHWSRVNKSLLNALKANVDGL 8469
QY 82 ----- 81
Db 8470 LEGUTYVFRCAENAAAGKFPSPDPKTAHDPISPFPPIPRVTDTSSTTIELEWEPPA 8529
QY 82 ----- 81
Db 8530 FNGGGEIVGVFDKOLVGTNEWSRCTEKMIKVRQYTVKEIREGADYKLRVSAVNAAGSGP 8589
QY 82 ----- 81
Db 8590 PGETOPVTVAEPOEPPAVELDVSVKGGIQIMAGKTLRIPAVVTGRPVPTKVMTKBEGELD 8649
QY 82 ----- 81
Db 8650 KDRVVIDNVGKSELI IKDALRKDHGRYVITATNSCGSKFAAARVEFVDFVGPVLDLPV 8709
QY 82 ----- 81
Db 8710 VTNRKXCLLNWSDPEDDGGSEITGFIIERKDAKMTWQPIETERSKCDITGLLEGQEVK 8769
QY 82 ----- 81
Db 8770 FRVIAKNKFGCGPPVEIGPILAVDPLGPPTSPELTYTERTKSTITLDWKEPSNGSGPI 8829
QY 82 ----- 81
Db 8830 QGYIIIEKRHDXDPFERNVNRKLCPTTSFLVENLDEHOMYEFVRKAVNEIGESEPSPLNV 8889
QY 82 ----- 81
Db 8890 VIQDDEVPTIKLRLSVRGDTIKVKAGEPVHIPADVTGLPMPKIEWSKNETVIEKPTDAL 8949
QY 82 ----- 81
Db 8950 QITKEEVSREAKTELSIPKAVREDKGTVTASNRLSGSVFRNVHVEYVDRPSPPRLAV 9009
QY 82 ----- 81
Db 9010 TDIKAESCYLTWDAPLDNGGSEITHYVIDKRDAKRKAEEVNTAVKRYGIWKLIPN 9069
QY 82 ----- 81
Db 9070 GQYEFVRVANKYGISDECKSDKVVIQDPVRLPGPGPKVLARTKGSMLVSWTPPLDNG 9129
QY 82 ----- 81
Db 9130 GSPITGWLEKREEGSPYMSRVSRAPITKVGLKGVFNVRLLLEGVKYQBAMAINAAGI 9189
QY 82 ----- 83
Db 9190 GPPSESPDEVAGDPIFPFGPPSPCEVKDKTKSSISLGHKPPAKDGGSPKIGYIVEMQEE 9249
QY 84 ----- 83
Db 9250 GTTDKRVNEPDKLITTCCECVVFNKLKELRYRFRVRAVNEAGESEPSDTTGEIPATDIOE 9309
QY 84 ----- 83
Db 9310 EPEVFIDIGAQCCLVCKAGSQIRIPAVIKGRPTPKSSWEPFGKAKAMKGVHDIPEDAQ 9369
QY 84 ----- 83
Db 9370 LETAENSSVIIIECKRSHTGKYSITAKNAGQKANTANCRVKVMVDPGPKDLKVSIDTRG 9429
QY 84 ----- 83
Db 9430 SCRLSMKMPDDGGDRIGKYGVIKRTIDGKATKVNDCGSTTFVVPDLLEQQYFFRVR 9489
QY 84 ----- 83

Db 9490 AENRFGIGPPVETIQTARTDPIYPPDPPIKIKGLITKNTVHLSWKPKNDGGSPVTHY 9549
Qy 84 ----- 83
Db 9550 IVECLAMDPTGCKEAMQCNKRDVEELQFTVEDLVEGGEYEFVRKAVNAAGVSKPSATV 9609
Qy 84 ----- 83
Db 9610 GPCDQRPDMPPSIDLKEFMEVEBGTNNIVAKIKGVPPFTLTWFKAPPKPDNKEPVLV 9669
Qy 84 ----- 83
Db 9670 DTHVNKLVDVDTCTLVIPQSRSDTGLYITAVNNLGTASKEMRLNVLGRPPVGPPIKF 9729
Qy 84 ----- 83
Db 9730 ESVSADQMTLSWFPKODGSKITNYVIEKREANRKTWVHSSEPKECTYTIIPKLEGE 9789
Qy 84 ----- 88
Db 9790 YVFRIMAQNKYIGIPLDSEPETARNLFSVPCAPDKPTVSSVTRNSMTVNWEEPEYDGS 9849
Qy 89 ----- 88
Db 9850 PVTGYWLEMKDTSKRKRVNRDPIKAMTLGVSVKVTGLIEGSDYQFRVYAINAAGVGA 9909
Qy 89 ----- 88
Db 9910 SLPSPDAPARDPIAPPPPPKVTDMTKSSADLEWSPLKDGSKVTGYIIVEYKEEGKEE 9969
Qy 89 ----- 88
Db 9970 WEKGDKVEVRGKLVVTGLKEGAFYKFRVSAVNIAGIGEPGEVTDVIEMKDRLVSPDLQL 10029
Qy 89 ----- 88
Db 10030 DASVRDRIWVHAGVIRIIAYVSGKPPPTVWNMNERTLPOBATTIETTAISSMVINKCQ 10089
Qy 89 ----- PEDD--- 92
Db 10090 RSHQGVYSLAKNEAGERKKTIIVDVDPGVGTPFLAHLNLTNESCCLKTWFSPEDDGS 10149
Qy 93 ----- 92
Db 10150 PITNVIEKESDRRAWTPTVYTRQNA TVQGLIQKAYFERIAAENSIGMPFETSE 10209
Qy 93 ----- 92
Db 10210 ALVIREPITVPERPELVEKVTQNTVLTWNPVKYDGGSEIINYVLESRLIGTEKPHKV 10269
Qy 93 ----- 92
Db 10270 TNDNLSRKYTVKGLKEGDTYEVRSVAVNIVGGKPSFCTKPICTCKDELAPPTLHDFRD 10329
Qy 93 ----- 92
Db 10330 KLTVRGEAFALTGRYSGKPKVSWFKDEADVLEDDRTHIKTTPATLALEKIKARSDS 10389
Qy 93 ----- 92
Db 10390 GKYCVVVENSTGSRKGCQVNVVDRPVPVGVSVFDEVTQYVVISWKPPDDGSKITN 10449
Qy 93 ----- 92
Db 10450 YIIEKKEVKDVMVPTASAKTTCKVSKLLEKDYIFRIHAENLYGISDPLVSDSMKAK 10509
Qy 93 ----- 92
Db 10510 DRFRVPADPQPIVTEVKDSALVTWNKPHDGGKPTNYILEKRETMSKRWARVTXDPH 10569
Qy 93 ----- 92
Db 10570 PYTKFRVPDLLEGQCFRVSARENIGDIPSPSPKPVFAKODIAPKPSPPVNPDEADTTTC 10629

Qy 93 ----- 92
Db 10630 NSVDLTWQPRHDGSKILGYIVEYQKVGDEEWRANHTPESCETKYKVTGLRQOTYK 10689
Qy 93 ----- TDPGH--- 97
Db 10690 FRVLAVNAAGESDPAHVPEPVLVKORLEPILLDANMAREQHIKYGDTLRLSAIIGVP 10749
Qy 98 ----- 97
Db 10750 FPKVTWKEDRDAPTAKRIDVTPVSGSKLEIRNAAHEDGGIYSLTVENPAGSKTVSVKVLV 10809
Qy 98 ----- 97
Db 10810 LDKPGPPROLEVSEIRKDCYLTKPEPLDGGSVITNYVVERRDVASAQSPLSATSKKK 10869
Qy 98 ----- 97
Db 10870 SHFAKHLNEGNYLFRVAAENYQGRFFVETPKPIKALDPLHPPGPKDLHHVDVDKTEV 10929
Qy 98 ----- 97
Db 10930 SLVWKPDRDGGSPITGYLVEYQEEGTQDWIKFVTNLECVVTGLQOGKTYRFRVKAEN 10989
Qy 98 ----- 97
Db 10990 IVGLGLPDTTPIECOEKLVPPSVELDVKLEGLVVKAGTVTFPAIIRGVPPVPTAKWTT 11049
Qy 98 ----- 97
Db 11050 DGSEIKTEHYTVETDNFSSVLTIKNCLRDTGEYQITVSNNAAGSKTVAVHLTVLDVPGP 11109
Qy 98 ----- 97
Db 11170 HLQKGEYVFRVAENKIGVPLDSTPTVAKHKFSPSPGKPVVTDITENAAVTSWTL 11229
Qy 98 ----- 97
Db 11230 PKSDGSGPITGYMERREVTGKVRVNNKPIADLKFRVTGLYEGNTYEFVRFAENLAGLS 11289
Qy 98 ----- ADLV--- 101
Db 11290 KPSSSDPIKACRPIKPPGPPINPKLKDRETDALVMTKPLSDGSGPILGYVVECOKPG 11349
Qy 102 ----- 101
Db 11350 TAOWNRINKOBLIRQCAFVRVPLGIEGNEYFRIKAANIVGEGEPRELAESVIAKDILHPP 11409
Qy 102 ----- 101
Db 11410 EVELDVTCDVITVRVGTIRILARVGRPEPDTITWKEGKVLVREKRVLDIQLPRVEL 11469
Qy 102 ----- 101
Db 11470 QIKEAVRADHGKYIISAKNSSGHAQSAIVNVLDROPQCNLKVNTVTKENCTISWENPL 11529
Qy 102 ----- 101
Db 11530 DNGSEITNFIVEYRKNQKQWSIVASDVTKRLIKANLLANNEYFRVCAENKVGVGPTI 11589
Qy 102 ----- 101
Db 11590 ETKTPIALNIDRPGEPENLHIAKGTFFVYLKWRRPDYDGGSPNLSYHVERRLKGSD 11649
Qy 102 ----- 101
Db 11650 WERVHKGSIKETHVMVDRCVENQIYFRVQTKNEGGSBWVKEVVVVKEDLQKPVLDLK 11709

Qy 102 ----- 101
Db 11710 LSGVLTAKAGDIRLEAGYRGKPFPEVAMTKOKDATDLTRSPRVKIDTRADSSKFLTKA 11769
Qy 102 ----- 101
Db 11770 KRSDBGKVVVATNTAGSFVAVATVNVLDKPGVPRNLKIVDVSSDRCTVCWPPDDGGC 11829
Qy 102 ----- 101
Db 11830 EIQNYILEKCKRMWSTYSATVLTGCTTVTRLIEGNEYIFRVRAENKIGTGPTTESKP 11889
Qy 102 ----- 101
Db 11890 VIAKTYDKGRDPDPPEVTKVSKSEMTVVMNPPEYDGGKSIITGYFLEKKEKHSTRWVPV 11949
Qy 102 ----- 101
Db 11950 KSAIPERRMKVQNLPLDPHEYQFRVKAENEIGIGESLPSRPPVAKDPIEPGPPTNFRVV 12009
Qy 102 ----- 101
Db 12010 DTTKHSITLWGKPVYDGGAPIIGYVVMRPKIADASPDGWMKRCNAAQLVRKEFTVTS 12069
Qy 102 ----- 101
Db 12070 LDENQEYFRVCAQNVGIGRPAELKEAIKPKBILEPPEIDLDASMRKLVIVRAGCPRL 12129
Qy 102 ----- 101
Db 12130 FAIVRGRPAKVTRKVGIDNVVRKGQVDLVDWAFVLPINSTRDDSGKSLTLVNPAGE 12189
Qy 102 ----- 101
Db 12190 KAVFNVVRVLDTPGVSUCLKVSDVTKTSCHVSWAPPENDGGSQVTHVIVEKREADRKTWS 12249
Qy 102 ----- 101
Db 12250 TVTPEVKKTSFHVTLNVPNGNEYFRVTAVNEYGPGVTDVPKPVLASDPLSEDPKPLE 12309
Qy 102 ----- 101
Db 12310 VTEMKNSATLAWLPPRLDGGAKIDGYITSYREEEQPADRWTEYSVVKDLSLVVTGLKEG 12369
Qy 102 ----- 101
Db 12370 KKYFRVAARNAGVSLPREAEGVYEAKEQLLPKILMPEQITIKAGKCLRIEAHVYCKP 12429
Qy 102 ----- 101
Db 12430 HPTCKWKGEDEVVTSSHVAHKADSSILIKDVTRKDSGYSLTAENSSGTDQKIKV 12489
Qy 102 ----- 101
Db 12490 VMDAPGPPQPPFDISDIDACSLSWHI PLEDGGSNITNIVIEKCDVSRGDMVTLASV 12549
Qy 102 ----- 101
Db 12550 TKTSRVRGLIPQGEYIFRVRAENRFGISEPLTSPQWQAPFPFGVPSEPKNARVTKNKD 12609
Qy 102 ----- 101
Db 12610 CIFVAMDRPDSGGSPIIIGYLIERKERNLSLLWVKANDTLVRSTEYPCAGLVEGLEYSFRI 12669
Qy 102 ----- 101
Db 12670 YALNKGSSPPSKPTEYVTARMPDPPGKPEVIDVTKSTVSLIWARPKHGGSKIIGYFV 12729
Qy 102 ----- 101
Db 12730 EACKLPDGKWRNTAPHQIPOEYATATGLEEKAQYQFRATARTAVNISPPSESPDPVTI 12789
Qy 102 ----- 101

Db 12790 LAENVPPRIDLSVAMKSLLTVKAGTNVCLDATVFGKPMPTVSWKKDGTLLKPAEGIKWAM 12849
Qy 102 ----- 101
Db 12850 QRNCLTLEFSEVNRKSDGYTITAENSNGSGKSATIKLVLDKPPASVKINKWYSRAM 12909
Qy 102 ----- 101
Db 12910 LSWEPPELDDGGSEITNIVDKRETSRPNWAQVSATVPITSCSVEKLIIEGHEYQFRICAEN 12969
Qy 102 ----- 101
Db 12970 KYGVCDPVFTPEPAIAKNPYDPPGRCDPEVINSITKDHMTVSWKPPADGGSPITGYLLEK 13029
Qy 102 ----- 101
Db 13030 RETQAVNWKVNRKPIIERTLKATGLOEGTEYEFPRVTAINKAGPKGKPSDASKAAYARDPO 13089
Qy 102 ----- 101
Db 13090 YPPAPPAPPKVYDTRRSSVSLSWGKPAYDGGSPIIGYLVEVKRADSDNWRNCLPNQLQK 13149
Qy 102 ----- 101
Db 13150 TRFEVTGLMEDTQYQFRVYAVNVKIGYSDPDPKHYPKIDILIPPEGELDADLRKTLILR 13209
Qy 102 ----- 101
Db 13210 AGVTMRLYVPVKGRPPPKITWSKPNVNRDRIGLDIKSTDFTDLRCENNVNKYDAGKYIL 13269
Qy 102 ----- 101
Db 13270 TLENSCGKKEYTIVVKVLDTPGPPVNVTVKEISKDSAYVWTEPPIIDGGSPINIYVQKR 13329
Qy 102 ----- 101
Db 13330 DAERKSWSTVTTECSKTSFRVANLEEGKSYFPRVFAENYIGIDPGETRDVAKASQTPGP 13389
Qy 102 ----- 101
Db 13390 VVDLKVRSVKSSCSIGHKKPHSDGSRIGYVVDFTLEENKQWRVMKSLSLQYSAKDLT 13449
Qy 102 ----- 101
Db 13450 EGKEYTFRVSAENENGEGTPEITVVARDVVAPDLKGLPDLCLYAKENSNFRUKIPI 13509
Qy 102 ----- 101
Db 13510 KGPAPSVSWKXGEDPLATDTRVSVESSAVNTTLIVYDCKSDAGKYITIMKNVAGTKEG 13569
Qy 102 ----- 101
Db 13570 TISIKVVGKPGIPTGPIKFDEVTAEAMTLKMAPKDDGGSEITNYILEKSDSVNNKWVTC 13629
Qy 102 ----- 101
Db 13630 ASAVQKTTFRVTRLHEGMEYTFRVSAENKYGVGEGLKSEPIVARHPFDVPDAPPPNIVD 13689
Qy 102 ----- 101
Db 13690 VRHDSVSLTWDPKKTGGSPITGYHLEFKERNLSLLWKRANKTPIRMDFKVTGLTEGLEY 13749
Qy 102 ----- 101
Db 13750 EFRMAINLAGVKPSLPSEPVVALDPIPPGKPEVINITRNSVTLIWTEPKYDGGHKL 13809
Qy 102 ----- 101
Db 13810 GYIWEKRDLPKSWMKANHVNPPECATVTDLVEGGKYEFRIAKNTAGAISAPSESTET 13869
Qy 102 ----- 101

```

Db 13870 IICKOEYEAPTVLDPTIKDGLTIKAGDTIVLNAISILGKPLPKSSWSKAGKDIRPSDIT 13929
Qy 102 ----- 101
Db 13930 QITSTPTSSMLTIKYATRKDAGEYTTATNPFGTKVEHVKVTVLDVPGPPGVEISNVSA 13989
Qy 102 ----- 101
Db 13990 EKATLTWTPLEDGGSPIKSYILEKRETSRLLTWTVVSEDIQSCRHVATKLIQONEVIFRV 14049
Qy 102 ----- 101
Db 14050 SAVNHYKGEPVQSEPKVMVDRFPGPPGPEKPEVSNVTNTATVSWKRPVDDGSGSEITGY 14109
Qy 102 ----- 101
Db 14110 HVERREKSLRWRAIKTPVSDLRCKVTGLQEGSTYEFVRVSAENRAGIOPPSSEASDVLM 14169
Qy 102 ----- 101
Db 14170 KDAAYPPGPPSNPHVTDTTKKSASLANGKPHYDGLLEITGYVVEHOKVGDEAWIKDITGT 14229
Qy 102 -LYITRF----- 107
|_|_|_|
Db 14230 ALAITQFVWPDLOTKEKNFRISAINDAGVGEPVDPVEIVEREMAPDFELDAELRTL 14289
Qy 108 ----- 107
Db 14290 VVRAGLSIRIFVPIKGRPAPEVTWTKDNINLKNRANIENTESFTLLIIPECNRYDTGKFV 14349
Qy 108 ----- 107
Db 14350 MTIENPAGKSGFVNVVRVLDTPGVLNLRPTDITKDSVTLHMDPLIDGGSRTNIVIEK 14409
Qy 108 ----- 107
Db 14410 REATRKSYSTATTKCHKCTYKVTGLSGCEYFFVRVMAENEYEGICEPTETTEPVKASEAPS 14469
Qy 108 ----- 107
Db 14470 PPDSLNIMDITKSTVSLAWPKPKHDGSKITGYVIEAQRKSGDQWTHITTVKGLECVVRN 14529
Qy 108 ----- 107
Db 14530 LTEGEETTFQWMAVNSAGRSAPRESRPVIVKEQTMPELDLRGIYQKLVIAKAGDNIKVE 14589
Qy 108 ----- 107
Db 14590 IPVLRPKPTVMKKGDQILKQTORVNFETTATSTILNINECVRSDSGYPPLTARNIVGE 14649
Qy 108 ----- 107
Db 14650 VGDVITIQHDIPIGPPTGIKFEVSDVFTFSWDPDENGGVPISNVVEMRQDSTTW 14709
Qy 108 ----- 107
Db 14710 VELATTVIRTTYKATRLTTGLEQYQFRVKAQNRGVGPGIITSACIVANYPFKVPGPGTPQ 14769
Qy 108 ----- 107
Db 14770 VTAVTKDSMTISWHEPLSDGGSPILGYHVERKERNGILMOTVSKALVPGNIFKSSGLTDG 14829
Qy 108 ----- 107
Db 14830 IAYEFVRVIAENWAGSKSPSEPMALDIPDPPOKVPVPLNITRHTVTLKWAKPEYTGCF 14889
Qy 108 ----- 107
Db 14890 KITSYIVKEKRLPNGRWMLKANFSNILENEFTVSGLTDAAYEFVRVIAKNAAGAISSPPSEP 14949
Qy 108 ----- 107
Db 14950 SDAITCRDDVEAPKIKVDYKFDVTILKAGEAFRLLEADVSGRPPTMWSKDGKELECTA 15009

```

```

Qy 108 ----- 107
Db 15010 KLEIKIADFSTNLVNDKSTRDRSGAYTLTATNPGGFAKHIFNVKVLDRPGPPEGLAVTE 15069
Qy 108 ----- 107
Db 15070 VTSEKVLWSFPPLDDGAKIDHYIVQKRETSRLAWTNVASEVQVTKLVTKLLKGNEYI 15129
Qy 108 ----- 107
Db 15130 FRVMAVNYGVGEPLSEPLESEPLAVNPYPGPPDPKPNPEVTTITKDSMVVCGHPDSDGSEI 15189
Qy 108 ----- 107
Db 15190 INYIVERRDKAGORWIKCNKKTLTDLRYKVSGLTEGHEBYEFRIEMAENAGISAPSTSPF 15249
Qy 108 ----- 107
Db 15250 YKACDTVPKPPGPNRVLDTSRSSISIAMNKPIYDGGSEITGYMVEIALPEDEMQIVT 15309
Qy 108 ----- 107
Db 15310 PPAGLKATSYTITGLTENQEKIRIYANNSEGLGEPALVPGTPKAEDRMLPPEIBLADL 15369
Qy 108 ----- 107
Db 15370 RKVVTIRACCTLRLFPVPIKGRPAPEVKWARDHGESLDSKASIESSTSYTLIIVGNVNRFD 15429
Qy 108 ----- 107
Db 15430 GKYIILTVENSGSKSAFVNVRLDTPGPPQDLKVKEVTKTSVTLTWDPPLDGGSKIKNY 15489
Qy 108 ----- 107
Db 15490 IVEKRESTRKAYSTVATNCHKTSMKVQDLQEGCSYVFRVLAENEYIGLPAETAESVKAS 15549
Qy 108 ----- 107
Db 15550 ERPLPGKITLMDVTRNSVLSWEKPEHDGSRILGIYIVEMOTKSGDKWATCATVKVTEA 15609
Qy 108 ----- 107
Db 15610 TITGLIQEEYSFRVSAQNEKISDPRQLSVPIAKDLVIPPAPKLLFNTFTVLAGEDLK 15669
Qy 108 ----- 107
Db 15670 VDVPPFGRPTPAVTHKONVPLKQTRVNAESTENNSLLTIKDACREDVGHYVVKLTNSA 15729
Qy 108 ----- 107
Db 15730 GEAIETLNVILDKPGPPTGPVKMDEVADTSITLSWGPYPKYDGGSSINNYIVEKRDTSIT 15789
Qy 108 ----- 107
Db 15790 TWOIVSATVARTTIKACRLKTGCEYQFRIAENRYKSYLNSSEPTVAQYPPFKVPGPGT 15849
Qy 108 ----- 107
Db 15850 PVVTLSSRDSMEVQMNPEISDGGSRVIGYHLERKERNISILMWKLNKTPIPOTKFKTTGLE 15909
Qy 108 ----- 107
Db 15910 EGVEYEFVRVSAENIVGIGKPSKVSECVARDCPPGRPEAIIVTRNSVTLQWKKPTYDG 15969
Qy 108 ----- 107
Db 15970 GSKITGYIVEKKELPEGRWMAKSFNTNIDTHFEVTGLVEDHRYEFVRVIAENAGVFPSEPS 16029
Qy 108 ----- 107
Db 16030 ESTGAITARDEVDPPIRISMDPKYKDTIVVHAGESFKVDADYIGKPIPTIQIWKQOELSN 16089

```


Qy	108	-----	107
Db	16090	TARLEIKSTDFATSLSVKDAVRVDSGNVILKAKNVAGERSVTNVKVLDRPPEGPVVI	16149
Qy	108	-----	107
Db	16150	SGVTAEXCTLAWKPLODGGSDIINYIVERRETSRLVMTVVVDANVOTLSCKVTKLLEGNE	16209
Qy	108	-----	107
Db	16210	YTFRIMAVNKYGVEPLESEBPVAKNPFVVPDAPKAPEVITVTKDSMIVWERPASDGG	16269
Qy	108	-----	107
Db	16270	EILGVLEKRDKEGIRWTRCHKRLIGELRLRVLTGLIENHDYEFVSAENAAGLSEPPSP	16329
Qy	108	-----	107
Db	16330	AYQACDPIYKPGPPNPKVIDITRSSVFLSWSPKIYDGGCEIOGYIVEKCDVSGEWTM	16389
Qy	108	-----	DL 109
Db	16390	CTPTGINKTNIEVEKLEKHEYNFRICAINKAGVGEHADVPGPPIVEEKLEAPDIDL	16449
Qy	110	EL-----	111
Db	16450	ELRKIINIRAGSLRLFVPIKGRPTPEVKWGVDEIRDAI1DVTSSFTSLVLDNVNRY	16509
Qy	112	-----	111
Db	16510	DSGKYLTLSENSGTSKSAFVTVRVLDTPSPVNLKVTBITKDSVITWEPPLDGGSKIK	16569
Qy	112	-----	111
Db	16570	NYIVEKREATRKSAAVNTCHKNWKIDQLEGCSYFRVTAENEYIGLPAQTADPIK	16629
Qy	112	-----	111
Db	16630	VAEVQPPGKITVDVTRNSVLSWTKPEHDGSKI1QIYIVEMQAKHSEKSECARVKS	16689
Qy	112	-----	111
Db	16690	QAVITNLTOGEYLFVRVAVNEKGRSDPRSLAVPIVAKDLVIEPDVKPAFSSYSVQVGD	16749
Qy	112	-----	111
Db	16750	LKIEVPISGRPKETITWTKDGLPLKQTRINVTDSLDTLTLSEIKETHKDDGGQYGITVAN	16809
Qy	112	-----	111
Db	16810	VVGQKTASIEIVTLDKPDPKPGVPKFDVSAESITLSWNPPLYTGCCQITNYIVOKRDTT	16869
Qy	112	-----	111
Db	16870	TTVWDVVSATVARTTLKVTKLKTGTEYQFRIFAENRYGOSFALESPIVAQVPYKEPGPP	16929
Qy	112	-----	111
Db	16930	GTPFATAISKDSMWIQWHEPVNNGSPVIGYHLERKERNISILWTKVNTKTI1IHDTOFKAON	16989
Qy	112	-----	111
Db	16990	LEEGIEYFRVYAENIVGVKASKNSECYVARPCDPPGTPPEIMVKRNEITLOWTKPYV	17049
Qy	112	-----	111
Db	17050	DGSGMITGYIVEKRDLPDGRMKASFTTNVIETQFTVSGLTEDQRYEYFRVIAKNAAGAIK	17109
Qy	112	-----	111
Db	17110	PSDSTGPITAKDEVELPRISMDPKFRDTIVVNAGETFRLEADVHGKPLPTIEWLRGDEI	17169
Qy	112	-----	PDGNR 116

Db	17170	ESARCEIKNTDFKALLIVKDAIRIDGGQYILRASNVAGSKSPVNVKVLDRPPEGPV	17229
Qy	117	QVRGVT-----OLGG-----	126
Db	17230	QVTGVTSEKSLTSPPLQDGGSDISHYVVEKRETSRLAWTVVASEVVTNSLKVTKLLEG	17289
Qy	127	-----	126
Db	17290	NEVYFRIMAVNKYGVEPLESAPVLMKPPFVLPGPPKSLLEVNTIAKDSMTVCWNRPSD	17349
Qy	127	-----	126
Db	17350	GSEIIIGYIVEKRDGRSGIRWIKCNKRITDLRLRVLTGLTEHDEYEFVSAENAAGVCEP	17409
Qy	127	-----	130
Db	17410	ATVYKACDPVFKPGPPTNAHIVDTTKNSITLAWGKPIYDGGSEILGYVVEICKADEEW	17469
Qy	131	-----	130
Db	17470	QIVTPOTGLRVTRFEISKLTEHDEYKIRVCALNKVGLGEATSVPGTVKPEDKLEAPDL	17529
Qy	131	-----	133
Db	17530	DSELKGIIVRAGGSARIHIHPFKGRPTPEITWSREEGEFTDKVQIEKGVNYTOLSDNCD	17589
Qy	134	-----	133
Db	17590	RNDAGKIILKLENSGSKSAFVTVKVLDTGPPONLAVKEVKDSAFVMEPPIIDGGAK	17649
Qy	134	-----	133
Db	17650	VKNYVIDKRESTRKAVANVSSCKSTSKVENLTEGAIYFRVMAENEFGVGPVETDA	17709
Qy	134	-----	133
Db	17710	VKAAEPSPGKVTLTDVTSQTSASLMWEKPEHDGGRVLGVYVEMOPKGTSEKSIVAESK	17769
Qy	134	-----	133
Db	17770	VCNVWTLSSGGEYQPRVKAINEKSGSDPRVLGVPVIAKDITIOPSLKLPFNTYSIOAG	17829
Qy	134	-----	133
Db	17830	EDLKIEIPVIGRPRPNISWVGDEPLKQTRVNVVEATATSTVLHIKEGNKDDFKYTVTA	17889
Qy	134	-----	133
Db	17890	TNSAGTATENLSVIVLEKPGPPVGPVRFDVSDAFVVISWEPAYTGGCISNVIVEKRD	17949
Qy	134	-----	133
Db	17950	TTTTWHVVSATVARTTIKITKLKTGTEYQFRIFAENRYGKSAPLDSKAVIVQYPFKEPG	18009
Qy	134	-----	133
Db	18010	PPGTPFTVTSISKDQMLQWHEPVNDGGTKIIGYHLEOKEKNSILWKLNTKPTIQTKEPT	18069
Qy	134	-----	133
Db	18070	TGLEGLEVEYFKVSAENIVGIGKPSKVSFCFVARPCDPPCPPEAIVITRNNVTLKWKPP	18129
Qy	134	-----	133
Db	18130	AYDGSKITGYIVEKKDLPDGRMKASFTNVLETEFTVSGLVEDQRYEYFRVIAARNAAGNF	18189
Qy	134	-----	133
Db	18190	SEPSDSGAITARDEIDAPNASLDPKYKDVIVVHAGETFRVLEADIRGKPIPDVWWSKOKG	18249
Qy	134	-----	133

Db 18250 ELEETAARMEIKSTIQKTTLLVVKDCIRTDGGQYILKLSNVGTSKSIPIITVKVLDPRGPPE 18309
Qy 134 ----- 133
Db 18310 GPLKVTGTAECYLAWNPPLQDGGANISHVIIKRETSRLSWTQVSTEQALNYKVTKL 18369
Qy 134 ----- 133
Db 18370 LPGNEYIFRVMAVKNYKIGEPLESPVTACNPKPKPPSPSTPEVSAITKDSMVVTWARPV 18429
Qy 134 ----- 133
Db 18430 DDGGTEIEGYILLEKXDKGVRWTKCNKKTLLDRLRLVTGLTEGHSYEFRAAENAGYGE 18489
Qy 134 ----- 133
Db 18490 PSEPSVFRACDALYPPGPPSPKVTDTSRSSVSLAWSKPIYDGGAPVKGYVVEVKEAAA 18549
Qy 134 ----- 133
Db 18550 DEWTTCTPPTGLOGKQFTVTKLKENTEYNFRICAINSEGVGEPAITPGSVVAQERIEPPE 18609
Qy 134 ----- 133
Db 18610 IELDADLRKVVVLRASATLRLFTVIKGRPEPEVKWEKABGILTDRAQIEVTSSTFMLVID 18669
Qy 134 ----- 133
Db 18670 NVTRFDGRYNLTLNENSGSKTAFVNVRLDPSAPVNLTIREVKKDSVTLSEWEPPLIDG 18729
Qy 134 ----- 133
Db 18730 GAKITNVIIVEKRETRKAYATITNNCTKTTFRIENLOEGCSYFFRVLASNEYIGLPAET 18789
Qy 134 ----- 133
Db 18790 TEPVKYSEPLPPGRVTLVDVTRNTATIKWEKPESDGSKITGYVVMOTKGSEKSTCT 18849
Qy 134 ----- 133
Db 18850 QVKTLEATISGLTAGEEYFVRAAVNEKGRSDPROLQVGPVIARDIEIKPSVELPFHTFNV 18909
Qy 134 ----- 133
Db 18910 KAREQLKIDVPFKRPOATVNRKDGQTLKETTRVNVSSKVTLSLSIKEASKEDVGTVE 18969
Qy 134 ----- 133
Db 18970 LCYSNSAGSITVPIITIIVLDPRGPPGPIRIDEVSCDSITISMNPPEDGQCISNYIVEK 19029
Qy 134 ----- 133
Db 19030 KETTSTTHIVSOAVARTSIKIVRLTTGSEYQFRVCAENRYGKSSYESSAVVAEYFPSP 19089
Qy 134 ----- 133
Db 19090 PGPGTPKVYHATKSTMLVTWQVPVNDGGSRVIGYHLEYKERSILWSKANKILIAQTOM 19149
Qy 134 ----- 133
Db 19150 KVSGLDEGLMEYRVAENIAGIGKSKCEPVPAEDCDPPGQPEVNTITRKSLSLKWS 19209
Qy 134 ----- 133
Db 19210 KPHYDGGAKITGYIVERRELDPGRWLKCNYNIOETVFEVTELTEDORYEFVRFARNAAD 19269
Qy 134 ----- 133
Db 19270 SVSEPESTGPIIVKDDVEPPRVMDVKFRDVIIVVKGAEVLKINADIAGRPLPVISWAKD 19329
Qy 134 ----- 133
Db 19330 GIEIEERARTEIISTDNHTLLTVKDCIRRDGTQYVTLKQNVAGTRSVAVNCKVLDKPGPP 19389

Qy 134 ----- 133
Db 19390 AGPLEINGLTAEKCSLSWGRPOBDGGADIDYIIVEKRETSHLAWTICEGELOMTSCKVTX 19449
Qy 134 ----- 133
Db 19450 LLKONEYIFRVTVGNKYGVEPLESVAIKALDPFTVPSPPTSLEIITSVTKESMTLCHSRP 19509
Qy 134 ----- 133
Db 19510 ESDGGSEISGYIIERREKNSLRWVRVNNKPVYDLRVKSTGLREGCEYEYRVYAENAAGLS 19569
Qy 134 ----- 133
Db 19570 LPSETSPILRAEDPVFLPSPSPKPIVDSGKTTITIAWVKPLPDGGAPITGYTVVEYKXSD 19629
Qy 134 ----- 133
Db 19630 DTDWKTISIQLRGTEYTIISGLTTGAEVYFRVKSVMKVGASDPDSSDPOIAKEREEBPLF 19689
Qy 134 ----- 133
Db 19690 DIDSEMRKTLIVKAGASFTMTVPGRGPVNVLMWPKDPTDLTRAYVDTTDSRTSLTIEN 19749
Qy 134 ----- 133
Db 19750 ANRDSGKYTLTIQNVLSAASLTIVVKVLDTPGPPTNITVQDVTKESAVLSWDVPENDGG 19809
Qy 134 ----- 133
Db 19810 APVKNYHIEKREASKAWSVTNNCNRLSYKVTNLQEGAIYYFRVSGENBFGVGI PAETK 19869
Qy 134 ----- 133
Db 19870 EGVKITEKPSPEKLGVTISKSDVSLTLWLPKPHDGSRIYHYVVEALEKGQKNWVKAV 19929
Qy 134 ----- 133
Db 19930 AKSTHHVVGLRENSYFFRVFAENQAGLSDPRELLLPVLKEOLEPPEIDMKNPSPSHV 19989
Qy 134 ----- 133
Db 19990 YVRAGSNLKVDPISGKPLPKVTLSDRGVPLKATMRFNTEITAENLTLINKESVTADAGR 20049
Qy 134 ----- 133
Db 20050 YEITAANSSTTKAFINIVVLDRPGPTGPVVISDITEESVTLKWEPPKYDGGSOVTNYI 20109
Qy 134 ----- 133
Db 20110 LLKRETSTAVWTEVSATVARTMMKMKLTGTEEYOFRIKAENRFGISDHIDSACVTVKLP 20169
Qy 134 ----- 133
Db 20170 YTTPOPPSTPWVNTRESITVGHHEPVNSGSAVVGHYHLEMKDRNSILWOKANKLVIRT 20229
Qy 134 ----- 133
Db 20230 THFKVTIISAGLIYEFRVYAENAGVKPSHPSEPLAIDACEPPRNVRITDISKNSVSL 20289
Qy 134 ----- 133
Db 20290 SMOQPAFDGSKITGYIVERDLDPGRWTKASFTNVTETQFIISGLTQNSQYEFVRFARN 20349
Qy 134 ----- 133
Db 20350 AVGSINPSEVGVPTICIDSYGVPIDLPLEYTEVVYKRYAGTSVKLRAGISCKPAPTIEW 20409
Qy 134 ----- 133
Db 20410 YKDDKELOTNALVCVENTTDLASILIKOADRNLNSGYELKLRNAMGSASATIRVOILDKP 20469

QY 134 -----CLIT 137
Db 20470 GPGGPIEFKTVABKITLLWRPPADGGAKITHYIIVEKRETSRVVMSVSEHLEECIIT 20529
QY 138 ----- 137
Db 20530 TTKIIGNEYIFRVRAVKNYKIGEPLSDSVAKNAFVTPGPGIPEVTKITNSMTVVW 20589
QY 138 ----- 137
Db 20590 SRPIADGSDISGYFLEKRDKSLGWFKVLKETIROTKVGTGLTENSQYRVCAVNA 20649
QY 138 ----- 137
Db 20650 GOGPFSEPFYKAADPIDPPGPAKIRIADSTKSSITLGSKPVYDGGSAVTGYVVEIR 20709
QY 138 ----- 137
Db 20710 QGEBEWTTVSTKGEVRTTEYVVSNLKPGVNYFRVSAVNCAGOGEPENNEVQAKDIL 20769
QY 138 ----- 137
Db 20770 EAPEIDLVALRTSVIAKAGEDVQLIPFKGRPPPTVTRKDEKILGSDARYSIENTDSS 20829
QY 138 ----- 137
Db 20830 SLLTIPQVTRNDTGKYLITENGVBEPKSTSVSVKVLDTPAACQKLOVKHVSRGVTLLW 20889
QY 138 ----- 137
Db 20890 DPPLIDGSPINIYVIEKRDATKRTWSVSHKCSSTSFKLDLSEKTPPFPRVLAENEIG 20949
QY 138 ----- 137
Db 20950 IGEPCETTEPVKAAEVAPIRDLMSKDSKTSVILSWTKPDFDGGSVITEYVVERKGGE 21009
QY 138 ----- 137
Db 21010 QTWSHAGISKTEIEVSQLEQSVLEPRVFAKNEKGLSDPVTGPTVKELIITPEVDLS 21069
QY 138 ----- 137
Db 21070 DIPQAQVTRIGHNVHLELPYKKGPKPSISWLKDGLPLKESEFVRFSKTENKITLSIKNA 21129
QY 138 ----- 137
Db 21130 KKEHGGKYTVILDNAVCRIVAITVITLGPSPKPGPIRFDKADSVILSWDPEDNGG 21189
QY 138 ----- 137
Db 21190 GEITCYSIEKRETSQTNWKMVCSSVARTTFKVPNLVKDAEYQFVRAENRYGUSQPLVSS 21249
QY 138 ----- 137
Db 21250 IIVAKHQFRIIPGPGKPIYVNTSDGMSLTWDAPVYDGGSEVTGFHVEKKERNISILQKV 21309
QY 138 ----- 137
Db 21310 NTSPIGREYRATGLVEGLDYQFRVYAENSAGLSPDPSKFTLAVSPVDPGPDYIDV 21369
QY 138 ----- 137
Db 21370 TRETTILKWNPLRDGSKIYGVYSIEKROGNERWRCNFTDVSECOYTVTGLSPGDRYEF 21429
QY 138 ----- 137
Db 21430 RIIRNAVGTISPSPSSGIIIMTRDENVPPIVEFGPEYFDGLIINKGESLRKALVQGRP 21489
QY 138 ----- 137
Db 21490 VPRVTWFKDGEIEKRMNMEITDVLGSLFVRDATRHRGVYTVAKNASGSAKAEIKV 21549
QY 138 ----- 137

Db 21550 KVQDTPGVVGPPIRFTNITGEKMTLWMDAPLNDGCAPIITHYIEKRETSRLAWALIEDKC 21609
QY 138 ----- 137
Db 21610 EAQSYTAIKLINGNEYQFRVSAVKNKFGVGRPLDSDPVVAQIQYTVDPAPGIPESNITGN 21669
QY 138 ----- 137
Db 21670 SITLTWARPESDGGSEIQOYLIERREKSTRVWVVISKRPISETRFKVGTGLTEGNEYEFH 21729
QY 138 ----- 137
Db 21730 VMAENAGVGPASGISRLIKCREPNPBPPTVVVKVTDTSKTTVSLWSKPVFDGMEII 21789
QY 138 ----- 137
Db 21790 GYIIEMCKADLGDWHKVNAAEACVKTTRYTVTLQAGEEYKFRVSAINGAGKGDSCVETGTI 21849
QY 138 ----- 137
Db 21850 KAVDLTAPELDIDANFKQTHVVRAGASIRLFIAYQGRPTTAVMSKPSDNSLSRADIIHT 21909
QY 138 ----- 137
Db 21910 TDSPTLTVENCNRNDAGKYTLTVENNSGSKSITFTVKVLDTPGPPGPIITFKDVTGRSAT 21969
QY 138 ----- 137
Db 21970 LMDAPLDGGARIHYYVVEKREASRRSQVISEKCTROIKFVNDLAEQVPPYFRVSAVN 22029
QY 138 ----- 137
Db 22030 EYGVCEPYEMPEPIVATEQPAPPRRLDVDTSSAVLAWLKPDDHGGSRITGYLLEMRQ 22089
QY 138 ----- 137
Db 22090 KGSDFWEAGHTKOLTFTVERLVEKTEYEFVKAKNDAGYSEPREAFSIIKEPOIEPT 22149
QY 138 ----- 137
Db 22150 ADLTGITNOLITCKAGSPFTIDVPISGRPAKVTKLEMLKEDTRVSIITTKDRTTLT 22209
QY 138 ----- 137
Db 22210 VKDSMRGDSGRYFLTLENTAGVKFTSVTVWIGRPGVPTGPIEVSSVSAESCVLSWGBPK 22269
QY 138 ----- 137
Db 22270 DGGGTEITNYIIVEKRESGTTAMQLVNSSVYKRTQIKVTHLTMYEYSFRVSGENRFGVSKP 22329
QY 138 ----- 137
Db 22330 LESAPIIAEHPFPVPSAPTRPEYVHVSAVANSIRWEEPYHDGSKTIIGYVVEKKERTILT 22389
QY 138 ----- 137
Db 22390 WVKENKVPCECNKYVTGLVEGLEYGQFRTYALNAAAGVSKASEASRPIMAONPVDAPGRPE 22449
QY 138 ----- 137
Db 22450 VTDVTRSTVSLIWSAPAYDGGSKVGYIIERKPVSEVGGGRWLKCNYYTIVSDNFFVTAL 22509
QY 138 ----- 137
Db 22510 SEGDTYEFVRLAKNAAGVISKSESTGPTVTCRDEYAPPAKELARLHLDLVTIRAGSDLV 22569
QY 138 ----- 137
Db 22570 LDAAGVKPEPKIITWTKGDKELDCEKSVLQYTGKATAVIKFCDRSDSGKYTLTVKNAS 22629
QY 138 ----- 137

Db 22630 GTKAVSMVKVLDSPGCKLTVSRVTOEKTCLAWSLPQEDGGAEITHYIVERRETSRLN 22689
Qy 138 ----- 137
Db 22690 WVIVEGECPTLSYVVTRLKNNKEYIFRVRANKYGGVPESEPIVARNSETIPSPGP 22749
Qy 138 ----- 137
Db 22750 BEVGTGKEHIIQWTKPESDGGNEISNLYVDKREKSLRWTRVNDYVYDTRLKVTSLM 22809
Qy 138 ----- 137
Db 22810 EGCDYQFRVTAVNAAGNSEPSSEASNFISCREPSYTPGPSAPRVVDTTKHSISLAWTKPM 22869
Qy 138 ----- 137
Db 22870 YDGGTDIVGVLEMOEKOTDQYRVHTNATIRNTEFTVPDLKMGOKYSFRVAANVKGMS 22929
Qy 138 ----- 137
Db 22930 EYSESIAEIPVERIEIPDELAADDLKKTVTIRAGASLRMLVSVSGRPPPVITWSKQIGID 22989
Qy 138 ----- 137
Db 22990 LASRAIIDTTESYSLLI VDKWRYDAGKYTIEAENOSGKSATVLKVYDTPGPCPSVKV 23049
Qy 138 ----- 137
Db 23050 KEVRSDSVTITWEIPTIDGAPVNNYIVEKREAAAMFAKTVTTKCSKTLYRISGLVEGTM 23109
Qy 138 ----- 137
Db 23110 YFPRVLPENIYGIGECETSDAVLSEVPLVPAKLEVVVDVTKSTVLAMEKPLYDGGSL 23169
Qy 138 ----- 137
Db 23170 TGVLEACKAGTERMMKVTLKPTVLEHTVTSLNEGEQYLFIRAQNEKGVSEPRETVTA 23229
Qy 138 ----- 137
Db 23230 VTQDLRVLPITDLSTMPQKTHVPAGRPVELVPIAGRPPPAASWFFAGSKLRESERV 23289
Qy 138 ----- 141
Db 23290 VETHTKVAKLTIRETIRDTGEYLELKNVTGTSETIKVILDKPGPTGPIKIDEIDA 23349
Qy 142 ----- 141
Db 23350 TSITISWEPPELDGAPLSGVVVEQDRAHPGWLPSVESVTRSTFKFTRLTEGNEYVFRV 23409
Qy 142 ----- 141
Db 23410 AATNRFIGSYLOSEVIECRSSIRIPGPPTLQIFDVSROGMLTWYPPEDDGGSQVTGY 23469
Qy 142 ----- 141
Db 23470 IVERKEVRADRWVRVNVPTWTRYRSTGLTEGLEYEHRVTAINARGSGKPSRSPKPIVA 23529
Qy 142 ----- 141
Db 23530 MDPIAPPGKQNPVRVTDTRTSVSLAWSVPEDEGSKVTGYLIEMQKVDOHEWTKCNTTP 23589
Qy 142 ----- 141
Db 23590 TKIREYTLTHLPOGAERYFRVLACNAGGPEAEVPGTVKVTMELYPDYELDERYQEGI 23649
Qy 142 ----- 141
Db 23650 FVRQGVIRLTIPIKGRFPPIKWTKEQDISKRAMIATSETHELVIKEADRGDSCTYD 23709
Qy 142 ----- 141
Db 23710 LVLENKCGKAVYIKVRVIGSPNSPEGPLEYDDIQVRSVRVSWRPPADGGADILGVILE 23769

Qy 142 ----- 141
Db 23770 RREVPKAAWYTIIDSRVGTSLVVKGLKENVEYHFRVSAENQFGISKPLKSEEPVTPKTPL 23829
Qy 142 ----- 141
Db 23830 NPPEPPSNPEVLDVTKSSVLSWSRPKDDGSGRVGTGYIERTKSTDKWVRHNKTOITT 23889
Qy 142 ----- 141
Db 23890 TMYTVTGLVPDAEYQFRI IAQNDVGLSETSPASEPVVCKDPFOKPSQSGELEILSISKDS 23949
Qy 142 ----- 141
Db 23950 VTLOWEKPECDGGKEILGYWVEYRQSGDSAWKSKNKERI KDKOPTIGLLEATEYEFVRP 24009
Qy 142 ----- 141
Db 24010 AENETGUSRPRRTAMSITKLTSGEAPGIRKEMKDVTTKLGEAAQLSCQIVGRPLPDIKW 24069
Qy 142 ----- 141
Db 24070 YRFGKELIOSRKYKMSDGRTHLTVMTEQDEGVVTCIATNEVEGETSSKULLOATP 24129
Qy 142 ----- 141
Db 24130 QFHPGYPLKEKYGAVGSTLRHLVMIYIGRPVPMWTFHGOGLQNSENITIENTEHTHL 24189
Qy 142 ----- 141
Db 24190 VMKNVQRKTHAGKYKVQVLSNVFGTVDAILDVEIODKPKPTGPIVIEALLKNSAVISWKP 24249
Qy 142 ----- 141
Db 24250 PADGGSWITNVVVEKCEKEGAEWQLVSSAISVTTCRIVNLTENAGYVFRVSAQNTFGI 24309
Qy 142 ----- 141
Db 24310 SDPLEVSSVWIIKSPFEKPGAPGKPTITAVTKDCVAVWPKPPADGGGAKIRNYVLEKREK 24369
Qy 142 ----- 141
Db 24370 KONKWI SVTTEIRETVPSVKNLIEGLEYEFRVKCENLGSESEISEPITPKSDVDPIQ 24429
Qy 142 ----- 141
Db 24430 APHFKEELNLRVYOSNATLVCKVTHPKPIVKWYROGKEIIADGLKYRIOEFKGGYHQ 24489
Qy 142 ----- 141
Db 24490 LIIASVTDATVYQVRATNOGGSVSGTASLEVEVPAKIHLPKLTLEGMVHALRGEVVS 24549
Qy 142 ----- 141
Db 24550 IKIPSGKPDPIVITWQKQDLIDNNGHYQVIVTRSFVSLVPNGVERKDAGFYVVCANR 24609
Qy 142 ----- 141
Db 24610 FGIDQKTVELDVADVPDPGRGVKVSVDVSRDSVNLTWBPSDGGSKITNYIVEKCATTA 24669
Qy 142 ----- 141
Db 24670 RMLRVGOARETRYVINLFGKTSYQFRVIAENKFGLSKPSSESEPTITKEDKTRAMNYDE 24729
Qy 142 ----- 141
Db 24730 EVDRETREVMTKASHSTKELYKYMIAEDLGRGEFGIVHRCVETSSKKTVMKFKVVKG 24789
Qy 142 ----- 141
Db 24790 TDQVLVCKEISILNIARHRNLIHLHESFESMEELVMIPEFISGLDIFERINTSAFELNBR 24849

```
QY 142 -----FDL----- 144
Db 24850 EIVSYHVQCEALQFLSHNIGHEDIRPENIIYOTRRSSTIKIEFCQARQLKPGDNFRL 24909
QY 145 ----- 144
Db 24910 LFTAPEYYAEVHQHDVVSTATDMWSLGLTVYVLLSGINPLAETNQOIIENIMNAEYTF 24969
QY 145 ----- 144
Db 24970 DEEAFKEISIEAMDFVDRLLVKKERKSRMTASEALQHPWLKOKIERVSTKVIRTLKHRRYY 25029
QY 145 ----- 144
Db 25030 HTLIKLDNVMVSAARISCGGAIRISQKGVSAKVAVASIEIGPVSGQIMHAVGEGGHVK 25089
QY 145 ----- 144
Db 25090 YVCKIENYDQSTQTVTYFVGRQLENSEKEYEITYEDGVAILYVKDITKLDDGTYRCKVVD 25149
QY 145 ----- 144
Db 25150 YGEDSSVAELFVKGVREYDYCYCRTMKKIKRRTDTWRLLRPPFTLPLYNKTYVGEN 25209
QY 145 ---GVTI 148
Db 25210 VRFGVTI 25216

RESULT 5
Q9EWA1 PRELIMINARY; PRT; 9507 AA.
ID Q9EWA1
AC Q9EWA1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pms2 protein.
GN PMS2.
OS Streptomyces natalensis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
ON NCBI_TaxID=68242;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20547809; PubMed=11094342;
RA Aparicio J.F., Foces R., Mendes M.V., Olivera N., Martin J.F.;
RT "A complex multienzyme system encoded by five polyketide synthase
RT genes is involved in the biosynthesis of the 26-membered polyene
RT macrolide pimarinin in Streptomyces natalensis.";
RL Chem. Biol. 7,895-905(2000).
RE EMBL; AJ278573; CAC20921.1; -.
DR HSSP; P25715; IMLA.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR002198; ADH short.
DR InterPro; IPR004410; Fabb.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000794; ketoacyl-synt.
DR InterPro; IPR001005; Myb_DNA_binding.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 6.
DR Pfam; PF00106; adh_short; 3.
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF00550; pp-binding; 6.
DR Pfam; TIGR00128; fabb; 6.
DR TIGRfams; TIGR00128; fabb; 6.
DR PROSITE; PS50075; ACP_DOMAIN; 6.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00037; MYB_I; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 5.
KW Phosphopantetheine.
SQ SEQUENCE 9507 AA; 994228 MW; 57BD80C8AD37E89F CRC64;
```

```
Query Match 34.2%; Score 264; DB 2; Length 9507;
Best Local Similarity 1.2%; Pred.No. 1.1;
Matches 98; Conservative 19; Mismatches 30; Indels 7866; Gaps 28;
QY 2 AGGILH-----LELLV----- 12
Db 86 SGGFLHDAAEAFDAPFGISPREALAMPDPOORLLESSHEAFERAGIDPNAVRGTRTGMFV 145
QY 13 -----AVGP----- 16
Db 146 GAMPQEVYRPDDDVQGFALTGTTTSVISGRLAYFFGAUGPAVTVDTCSSSLVALHLAA 205
QY 17 ----- 16
Db 206 HSLRQECSLAALAAAGVTVMSSPTTFVEFNRQGLSADGRCRSFADSADGTGSEGVLV 265
QY 17 -----DV 18
Db 266 LERLSEARNNGHEILAVIRGSANVDGASNGLTAPNGPSORRVIEQALYSARLSADEVDV 325
QY 19 FQAH----- 22
Db 326 VEAHGTGTTLGDVPEAQAALLATVGGGRDDDRPLLGSVKSNLGHTQAAAAGMAGVIKMWLA 385
QY 23 ----- 22
Db 386 MRHGVLPTLHVDAPSSHVDWTQGAVALLTTEHVPMQGPQRPAGVSFGLSCTNAHTIL 445
QY 23 -----QEDT----- 26
Db 446 EAPAAPSPGAGQEDTAPVAEPLAAGVPLVSGRTREALRAQAARLLEHLTSRPLGL 505
QY 27 ----- 26
Db 506 DVAYSLATARSGLHRAAFTTAAGPCGARAALTALADGTPAPGLVQDTARTSKLAFAG 565
QY 27 ----- 26
Db 566 QGSQRPQMGRELAAFPVFATALDEVLGHFDEGLERPLKDVLPAAEGTPEAALLDQTYA 625
QY 27 ----- 26
Db 626 QPALFAIEVALYRLAESFGITPDLFAGHSIGEIAAAHVAGVFTLADAAALVLARGRLMQA 685
QY 27 ----- 26
Db 686 LPEGAMVSLAEATEEEVLPLEDRPDKLSIAAVNGPAAVIAGEADDVTAAEHFDALGR 745
QY 27 ----- 26
Db 746 RTKRLRVSHAFSPLMPEMLAEFRFVAVSRILTPOAPVTPVVSGLTGLTATVEQLTSPDYWL 805
QY 27 ----- 26
Db 806 DHARHAVRFADALRLARRRRRLPRMGPDGVLSAQAQSCRGHRRRRRLRHARRVP 865
QY 27 ----- 26
Db 866 GRPEAETLTAALAGLHTRGVAVQWEPYFQGTGARRADLPTYAFQRRRYWPKSLPATGVDV 925
QY 27 ----- 26
Db 926 RAAGLGAHHPLTAAVSVANSOGLLTLGRLSRRTHPWADHAVRGTVLLPGTAFLELAV 985
QY 27 ----- 26
Db 986 RAGDEACGRVEELTAAPLVLPPEGGVQVQVMVSPDASGRRAVSVHSRDPGPEELPWT 1045
QY 27 ----- 26
Db 1046 QHAAGTLVAGBHHAGFDATVMPPADATPLDLGDFYARMADTFCGYGFLFOGLRAARSGD 1105
QY 27 ----- 26
```

Db	1106	DVTAEVLPDSGNA	SSAAGFLHPALLDAALHAAAFVDLGEDARGGLPFSQDVTLLH	1165
Qy	27	-----	-----	26
Db	1166	ASGASTVRVKLT	PDGDDAVALAVADTTGAPVASIGSLVLRTPDEQIDAAHSLVRDALFE	1225
Qy	27	-----	-----	26
Db	1226	LRMSEARQTAHEA	PASAAVLGPDLPFLGERALAPTDLTVTAVPGPDAAHPADPVVLAPVAG	1285
Qy	27	-----	ERYVLTNL	34
Db	1286	TTEDDVTSVHART	HAHVLERLRLTWLTDERYAASRLVFVTRGALATDARQASDPVAAAAMVG	1345
Qy	35	-----	-----	34
Db	1346	LVRAAQSEHPGR	FGLDLDPTGTTTEPATEPLRLALALSDEPQTAVRGTAVLTLARLARAQA	1405
Qy	35	-----	-----	34
Db	1406	PQAATEWDPECT	VTVLTGATGGLGRVLARHLVAERGVHRHLLLVSRRGADADGAGELMAELA	1465
Qy	35	-----	-----	34
Db	1466	AHGAQVTLAACD	LADRAAVRLLSGVPAEHPVTAVVHSAGSLDDGVIASLTPERISAVLR	1525
Qy	35	-----	-----	34
Db	1526	PKADAVWLH	HEATRDLDLAAFPWFSSLSGTVGAAQOGNYAANAFLDALAQLRRAOGLPG	1585
Qy	35	-----	-----	34
Db	1586	LSLANGPWAPT	AGTGTGLTDDDLRLARMGTPALTEOGIALFDAATADMAVLLPTRID	1645
Qy	35	-----	-----	34
Db	1646	LSVLRVQSEI	PPLMRGLIRTPARRSVVSASATAVTLVQOLSRLAETDRREVVLDLVAGV	1705
Qy	35	-----	-----	34
Db	1706	AAVLGHAGTAD	IDPRRPLRELGFDLSLTAVELNRLSAATGURTAATVIFDHTVDALAAH	1765
Qy	35	-----	-----	34
Db	1766	LLDGLMGEAAAA	ADAPRTAADDDPIVIGMSCRYPGGVASPEDLWRLVSEGTDAISGLP	1825
Qy	35	-----	-----	34
Db	1826	TDRGWDLALYD	PDPTGTSTYSRFGFLHTAADFPAPFGMSPREALATDSQORLLLEA	1885
Qy	35	-----	-----	34
Db	1886	SCEAVERAGID	PVLSRSGTGVFAGVMYNDYATVLSGOQFEGHOOGTAPSVASGRISYT	1945
Qy	35	-----	-----	34
Db	1946	LGLEGPAVTVD	TACSSSLVAMHWAMQALRAGECSLALAGGVTVMTPTGALLIEFSRQRGLS	2005
Qy	35	-----	-----	34
Db	2006	PDGRCKATSD	GADGVGWSEGVGVLVLERLSDARRNGHRVLAVVRSVAVNQDGASNGLTAP	2065
Qy	35	-----	-----	34
Db	2066	NGPAQORVIRQ	ALSAGLSATDIDAVEGHGTGTTILGDPLEAQALLAAYGQDREAGSPLLL	2125
Qy	35	-----	NIG-----	37
Db	2126	GSVKSNIIGHT	QAAAGVAGVIKVMAMRHCVLPRTLHADEPSSHVEWDSGAVRLLTETTPW	2185
Qy	38	-----	-----	37
Db	2186	POADRPRRAAV	SSFGSGTNAHLIVVEEPAVAERREEPALTPAVVPVWTLSGKSRAALRDO	2245
Qy	38	-----	-----	42
Db	2246	AARLSFLEER	PALDPAVALSLATRSAPDORAVVTGDRDELLRALADLAADRPGPALT	2305
Qy	43	-----	-----	46
Db	2306	EGEVGGAGKLA	VVFGSGQSQRPGAGRELAARPPVFAQALDEVTAALDPLHDLRPLKILPA	2365
Qy	47	-----	-----	46
Db	2366	PEGSPEALLDR	TEWTQPALFAIGVALHRLLTGTWGRIRPDVLLGHSIGEITAAHVAGVLSL	2425
Qy	47	-----	-----	46
Db	2426	PDAARLVIAGR	LMQALPAGGAMISLEATEDETAPILLAGRDHEVSLAAVNGPRSVVVAGE	2485
Qy	47	-----	-----	46
Db	2486	LAABEEIAAH	FADRGRTKRLTVSHAFSPLMEPALDGLREAAAEITYHAPDVTVISGLT	2545
Qy	47	-----	-----	46
Db	2546	GRPATEAEL	RSADYVYVHARGAVRFADTLURAAHDCGAGTFLBLGPDVSLSAAAQDALGDD	2605
Qy	47	-----	-----	46
Db	2606	AEAHIVPLLR	HGRGEBERCAATALARLHVGTADMPGYLAGTGARTIDLPVAFQHERYW	2665
Qy	47	-----	-----	46
Db	2666	PEQAPARLAQ	GAADPADAEELGWAVEGDATELAALLGLRDEQHASLYALLPALSSWROHR	2725
Qy	47	-----	-----	50
Db	2726	QEKALDSTRY	RVAMPVQATAAPVLDGTWLLVTADGIDSDEILDALRGHGAOFETLVLD	2785
Qy	51	-----	RVHL-----	54
Db	2786	DACRDAHLA	GLRAVVDQDRPLTGVLSLLPLADRSGAEGAGAGAGTGAQPSGLALS	2845
Qy	55	-----	-----	54
Db	2846	VLIQALADT	GLAAPLWTLTRGAVTTGADDPPLTHPLOAAVMGLGRVALEHPQLMSGILDL	2905
Qy	55	-----	-----	54
Db	2906	PAALDAPAT	QHLVLSALAAEGGEDQIAVRATGAPGRRLVRHPQAEPLPADAFRTGLIT	2965
Qy	55	-----	-----	54
Db	2966	GGTCALGAE	AAARWLARSGAEHLLTTSRRGPEAPGAELAAEIEELGARVTLAACDTADRA	3025
Qy	55	-----	-----	54
Db	3026	ALAAALLDSL	PDEHPLTGVVHAAGVQAAALADTPLADAAQMAAKLLGAHLLSLDGH	3085
Qy	55	-----	-----	54
Db	3086	LDPFVLVSS	IAGVWGSAGQSAYSAANNYLDALAEHRAARGLPATSVANGPWAEGMATHE	3145
Qy	55	-----	-----	54
Db	3146	AVTDELHKR	GLRFLAPATALAEELRAVVRDVTVTADIDWERYHPVTSTRPSALFDEL	3205
Qy	55	-----	-----	54
Db	3206	AEVQALIRPE	DESAAPEFAARLRGLDQDQGERLLGLVRAEAAIALGHDSADATAERRAF	3265
Qy	55	-----	-----	54
Db	3266	RDAGFDSLTA	VELRKRLAULTGLALPATLVFDYPTPVALARHLREQLLGTAOBSAAPVAV	3325

Qy	55	-----	54	Qy	75	-----	74
Db	3326	PSGGDEPIAIVGMSCRFPGGVRSRQLWDLVSEGVDAISDFPVNRGHWNTGLFHPDPNP	3385	Db	4406	DQAQILGLGHTAALEHPQRCGAGLVDLPELDRAGERLAAVLITGRSGEDOLAVRSSGVF	4465
Qy	55	-----	54	Qy	75	-----	74
Db	3386	GTTYSTOGFLHDAGEFDSFFGISPREALSMDPOORULLLETTWEAFEHAGIDPTTVHGT	3445	Db	4466	ARRVRAAPADPAPARPWTPRGTTLVGTGTTLAPHLARMLAQOQAEHIVLTSRRRGAAP	4525
Qy	55	-----	54	Qy	75	-----	74
Db	3446	TTGTFIGTYQBYGLGVEDGSAGHLVTGTSFVLSRGLAYLFLGLEGPATVDTACSSSLV	3505	Db	4526	GAELVQELAEQAEAVACDLTDRDAVQAQMLAALRAEGRTVTVVHTAVTIELAPLDE	4585
Qy	55	-----	54	Qy	75	-----	74
Db	3506	ALHLAQSLRNGESSLALAGGATVMTNPNPFVAFSRQALAGDGRCKAFSEGADGMTLAE	3565	Db	4586	TTLDDFAKVMDAKVIAGARHLDLGGDDDAFVLYSSTAGMWGSAHAAYVAGNAYLNALA	4645
Qy	55	-----	54	Qy	75	-----	74
Db	3566	GVGLVLERLSDAORNGHEILAVVRGSAINQDASNGLSAPNGSQQRVIRQALANSGLT	3625	Db	4646	EHRRGARATAVSWGIIWADDLKLGRVDPQGIIRRSGLVFMDPOLALTGLRQALDDEETHL	4705
Qy	55	-----	54	Qy	75	-----	74
Db	3626	PADIDAVEHGTGTALGDPIDIAQALLATYGPGRDPDPSPLLLGSKVSNIGHTQSAAGVAGV	3685	Db	4706	AVADVDMGRYYPVPTAARPTLRFEDLPEVTRQLAQESAEVTDGSEFATRLDGLAAVEODR	4765
Qy	55	-----	63	Qy	75	-----	74
Db	3686	IKMVLALRNGQLPRTLHADTPSSHIDWTPGTVSLLTPTDMPGERPRRCVSVSFGISGT	3745	Db	4766	LLLDLVRTEAATALGHASPDVLSQRAFRDVGFDLSLTAVDLRNRIRASVTGLALPSTMVFD	4825
Qy	64	-----	63	Qy	75	-----	74
Db	3746	NAHTLLEAPRPEATPAGVPVPPAGDAIPWLSARTPGALRAQAAQLAAHLDCGAPDALD	3805	Db	4826	YPNPLALVAFRESLAGSATGTAATTTHTAADEPIAIIIGMSCRYPGCVGSPEDLWRLVT	4885
Qy	64	-----	63	Qy	75	-----	74
Db	3806	VGHTLVAARTLFDHRAVVVGTDDASRAALDALATGSGAPGIVQGTADTDGRTVFVFPQ	3865	Db	4886	EGGATGEFFADRGWDAEGLYDPPDRAGHTYSTRGGFLHDAADFASFFGISPREALAM	4945
Qy	64	-----	67	Qy	75	-----	74
Db	3866	GSQWAGMARLLBESPVFAARLTCATALSEFVDSLLDLVRQADGAPTLDRVVDVVPAS	3925	Db	4946	DPQRLLETSWEARAGIDPATLRGSGTGTFIGASYQDYAFAGAGDGAEGHLITGTI	5005
Qy	68	-----	67	Qy	75	SSLLS-----VCG-----	82
Db	3926	FAVMSLAALWTSHTGTPDAVVGHSQGETAAAAVAGALSLEDAARVVVALRSQAIARGLAG	3985	Db	5006	SSVLSGRLSYTYGPEGPAVSLDTACSSSLVALHLACOSLRNGESSLALAGOVSIMSTPGA	5065
Qy	68	-----	67	Qy	83	-----	82
Db	3986	TGGMLSVPAPADVEQRLAAYEDLSIAAVNGPRSTVVSGATAPLDALQAEVLGVEDIRARR	4045	Db	5066	FVGFSPQAMAADGRCKAYSPOADGMSLAEGVLVLERLSDARRNGHQLAVIRGSAVN	5125
Qy	68	-----	67	Qy	83	-----	82
Db	4046	IAVDYASHSAQVERVRDELTVLEVPVRPAQVPFFSTVTGDLDTVMDAEYWFNLR	4105	Db	5126	QDGASGLTAPNGPSQQRVIRQALANAEVDPDGIDVIDGHGTGALGDPIDIAQALLATYG	5185
Qy	68	-----	67	Qy	83	-----	82
Db	4106	TVHFQPAIGELLAQGHFFIEVSSHPVLSMCIQATAEEAGAAAALVGLTRRDGTATDRFL	4165	Db	5186	QGRDPEHPLLGSVKSNIGHTQMASGVASVIMVMMRHAAPRSLHIDRPSHVDMSSG	5245
Qy	68	-----	67	Qy	83	-----	96
Db	4166	ASLAEAFVRCADADWSAVFAGTGARRVPLTYAFQREHLWAIPEPSDRPEADPADAEFW	4225	Db	5246	AIQLLTEPLPWPGTCHPRRAGVSSFGSLGTNVHTILEQAPEDRTBEDTDTAPARPADPTP	5305
Qy	68	-----	67	Qy	97	-----	102
Db	4226	TAVEEDVEALSLRLDRTSLAPVLPALSHWRKRRELSTVDSWRYRATWKPLTGLPAA	4285	Db	5306	VBVTLGRTEAGLRAQAGRLTHLADHPLEPLTDIAFSLATSRAALEHRAAVVTTDDPTL	5365
Qy	68	-----	74	Qy	103	-----	105
Db	4286	ALTGTWLVPAPEGADDTLTGVLAAHGAGTFTLVLDSDCADRATLTARLTGLDGAADLTG	4345	Db	5366	TRALTALRDATPDGGLITGRDPRGRGLAFLFTCGQSQRPCMGRRELYDRHPVYAEALDAVLA	5425
Qy	75	-----	74	Qy	106	RFDE-----	110
Db	4346	VVSLPADEASATHPTLSRAVTLTVALVQALGDAGIDAPLWCLUTQGAVSTGRADRLTRP	4405	Db	5426	RFDELDRLRLREILFADPGTPEAELDDTGTQTPALFALEVALFRLAESWGLURPDYVAGH	5485
				Qy	111	-----	114

Db 5486 SIGELAAAHACVLSLEDACTLVAAAGRLMAALPSSGGAMASVEATEDEAAAVALAPYEGRA 5545
|||
Qy 115 ----- 114
Db 5546 AVAAVNTPTSLVSGDEDAVHAAHFHGLGRTTKQLRVSHAFSPHMDAMLDQFTRAEE 5605
Qy 115 -----NRQVRG----- 120
Db 5606 GITFAPELPLVSLTUGELILADDLCSAPYWTQVRGTVRFADAVRHLEHGVHTFFPELG 5665
|||||
Qy 121 ----- 120
Db 5666 PDAVLGAVRESTGQTPATAVPALRRDRPEAPALTALARLHLHGVRVDWDVAFAGGA 5725
Qy 121 ----- 120
Db 5726 RRIDLPTYFORERYWEAAGAAAPADAAAEFWSAVERADLPISLSSLDLDDTL 5785
Qy 121 ----- 120
Db 5786 TAVVPALSSWRKSRERSTVDGWRVYRTTWKPLTGATAGHPAGTWLVLPADGDRAWTDA 5845
Qy 121 ----- 120
Db 5846 VAGALCADAVRVEVSTAQOQELAEERLSELAEHGEFAGVLSLLATAGEGAEDDATTGEG 5905
Qy 121 ----- 120
Db 5906 LLLTATALQALGDAGIDAPLWCVTRTAVAVDRAEHPARPAQAAVWGLGRVAALHPQRMG 5965
Qy 121 ----- 120
Db 5966 GLVDLPDELGDATLRLAAVLADSGEDQLAVRTTATFVRLAHHRAPAPAAATFRPTG 6025
Qy 121 -----VTQLG----- 125
Db 6026 TVLVTTGGTCALGHHVARWLAEAGAEHLLVSRRGTDAPCAGELAAEIAELGARVTVAACD 6085
:|:
Qy 126 ----- 125
Db 6086 TADRALAAVLAAPIDEHPLTAVFHTAGTVDDGTLTLTPEQFASVLRTKVATRNHET 6145
Qy 126 ----- 125
Db 6146 TREQDLSEFVLFSSVAGTLCAPQGGNYAAGNAFLDAFAAHRRAHGLPATSLAWGPWAETG 6205
Qy 126 ----- 125
Db 6206 MAADGTGIQDRVRRGGFTPMPPRLALTALRAIEHDATAALTLDIDWQRYAEVFTATRPS 6265
Qy 126 ----- 125
Db 6266 LFVADLPQLQVTPPAGTPEAALREPALRQLAGLSPAARPRFVLDLVRTOVAAVLHSG 6325
Qy 126 ----- 125
Db 6326 TSAIGERAFAFSLGFDLSLTTVELURNLTATTGTLKLPLATLVYDYPTTALADFLAELLCA 6385
Qy 126 ----- 125
Db 6386 LPESGAPAFVGRAADDPIVIVGMNCRFPFGIRSPEDLWQLSSGEDAISGFPADRGMWL 6445
Qy 126 ----- 125
Db 6446 DALARGASATLEGGLDGVGLFDAPFFGISPREALAMPQORLLLETSWEPERAGIDAT 6505
Qy 126 ----- 125
Db 6506 TLRGSRGTGVFGTNGQDYATLLRQGTGTTDIRGHVATGNTASVMSGRLSYAFGLEGPAYT 6565
Qy 126 -----GACSP-- 130
| |

Db 6566 VDTACSSALVALHMAAGALRSGECTTLALAGGVSVMSSPSDFTFTVQGLAPDGRCKPPA 6625
Qy 131 ----- 130
Db 6626 DAADGTSWSEGVGLLERLSDARRNGHDVWGIVRGTAVNODGASNGLTAPSGRAOORAI 6685
Qy 131 ----- 130
Db 6686 RQALADADLAPADVVEAHGTGTTGLDPIEAHALIAAYQDRPTDRPLLGLAVKSNLGH 6745
Qy 131 ----- 130
Db 6746 TORAAVAGVTIKVLLAMRHGVLPKLTHDPSAHVDMTACTVSLTTEQQEWPETGHHARRA 6805
Qy 131 ----- 130
Db 6806 GVSAGVSGTNAHVITEQAPDAESVAPEPTTVPLVPWPVSGKSEAALEAGIERVSALT 6865
Qy 131 ----- 130
Db 6866 GRSPVDVGHSLTTGTRAAFTHRAVLADGAGVRELARGVAHEADGKLAVLFSGOGAQRVGM 6925
Qy 131 ----- 130
Db 6926 GRELYVRFPVFAEALDGVLAHFDAGLRDVDMFGDAEGLDETGTQOPALFAIEVALFRLAES 6985
Qy 131 ----- 130
Db 6986 LGVRPDPFGHSGTGEIAAAHVAGVFSLADACALVAABARLMQALPAGGAMVAVQATAEEA 7045
Qy 131 ----- 130
Db 7046 AARLVDGVSIAAVNGPEAVVIAGEQREVLRIAKEFTGLGRKTRRLPVSHAFSPLMPEML 7105
Qy 131 ----- 130
Db 7106 EDFERRVAEGLSYEAPRIALVSNLTGELATEDLRSABEYWRHVHRETIVRFADGVSALAAEG 7165
Qy 131 ----- 130
Db 7166 ASVFEELGPDGVLTAQAHTLOGTATAVPALRKDMSEETALLTALAQLHVVGVDWMSGI 7225
Qy 131 ----- 130
Db 7226 FAGTGARRVDLPTYFQHQWYWDGAPAAEATTGAQDPVDAEFMDAVERADLSLSDSLE 7285
Qy 131 -----TWSCLI----- 136
Db 7286 LDDAALSALVPALSAMRRKRVVERSTVDGWRVYHTFKPLTGTAPOTLAGTWLALVPTGAAD 7345
|||
Qy 137 ----- 136
Db 7346 DEWTTAVLAALGDAAVIAEVDPPADRGAFAGLGLAERGTEFTGVVSSSSQSRTPSPSPD 7405
Qy 137 ----- 136
Db 7406 RGPTRRRWALAEAGIDAPLWCVTRDAVPVDRSESEVSPAQAAVWGLGRVAALDHPDRWG 7465
Qy 137 -----TEDT----- 140
Db 7466 GLVDVPOEHRRTAERLRAVLSGAGDEDOVALRPSGAFGRRLVRATDDTTPATPHQAGTV 7525
|||
Qy 141 ----- 140
Db 7526 LVVGPASGTGHCARWLAGOGAQLVLADRTAPDTAQLIGELDALGVPLTVLDCAPADGE 7585
Qy 141 ----- 140
Db 7586 ALLSALAALPEPLTAVIYADGPDADAYGTATTADPGASLAALYADVESLAABAAATGA 7645
Qy 141 ----- 140
Db 7646 RPLGAFVLFSSIAGTWAGGRGTAAAGAAQDALAGMLRARGLPATAVSWGWDATSGS 7705

141	--	140
7706	LAHLRANGLPMPDPARALTAARSLGTDTSVTADVVRWDEFAPATRRHSALFAELP	7765
141	--	140
7766	EAREALAESDGSAPOGTADAASGLRAELAARPDARTERHLLALVDRVAAVLGFADADAVP	7825
141	--	140
7826	SGQAFDTLGFDSLTAVDLRNQLAVATGLPATLVDFYPTADALAGYLRAELLGGKGKDA	7885
141	--	144
7886	EELPAVRTKDADTAEDPWIVGN [*] SCRYPGGISPEDLWLQVMGEADAIGGFPTDRGMWL	7945
141	--	144
7946	DRLLHGDRGRGRTVTQGQFLYDVADFPDGFPGFIAPREAMVMDPQORILLEATWEALER	8005
141	--	144
8006	TGIDPARLRGGDTVFICGGSGDYRPEAQGLHAQTQAQSALLSGRVAYHFLEGPVSUV	8065
141	--	148
8066	DTACSSSLVALHLAAQALNGECSTALTGGTVV	8098
141	--	148
RESULT 6		
Q09164		
ID	Q09164	PRELIMINARY; PRT; 15281 AA.
AC	Q09164;	
DT	01-NOV-1996 (TrEMBLrel. 01, Created)	
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)	
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)	
DE	Cyclosporin synthetase (CYSYN) (EC 6.-.-.-).	
EN	SIMA.	
OS	Tolypocladium inflatum.	
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	
OC	Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae;	
OX	Tolypocladium.	
OX	NCBI_TaxID=29910;	
RN	[1]_TaxID=29910;	
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	
RC	STRAIN=ATCC 34921;	
RX	MEDLINE=95094306; PubMed=8001164;	
RA	Weber G., Schoergendorfer K., Schneider-Scherzer E., Leitner E.;	
RT	"The peptide synthetase catalyzing cyclosporine production in	
RT	Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading	
RT	frame.";	
RL	Curr. Genet. 26:120-125(1994).	
CC	-I- FUNCTION: THE CONSTITUENT AMINO ACIDS OF CYCLOSPORINS ARE	
CC	ACTIVATED AS AMONACYL-ADENYLATES WITH PEPTIDE BONDS FORMED	
CC	THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER INTERMEDIATES.	
CC	ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.	
CC	-I- COPACTOR: CONTAINS 11 COVALENTLY BOUND PHOSPHOPANTETHEINES.	
CC	-I- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF CYCLOSPORINS.	
CC	-I- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT	
CC	COVALENT BINDING OF AMP TO THEIR SUBSTRATE.	
DR	EMBL; Z28383; CAA82227.1; -.	
DR	HSP; P14687; IAMU.	
DR	InterPro; IPR000873; AMP-bind.	
DR	InterPro; IPR000267; Asp/Glutamase.	
DR	InterPro; IPR001242; Condensatn.	
DR	InterPro; IPR003880; Plpantne attach.	
DR	InterPro; IPR000051; SAM_bind.	
DR	Pfam; PF00501; AMP-binding; 11.	
DR	Pfam; PF00688; Condensation; 13.	
DR	Pfam; PF00550; pp-binding; 11.	
DR	PRINTS; PR00154; AMPBINDING.	
DR	PRINTS; PR00139; ASGNGLNASE.	

QY	12	-----	11	QY	19	-----	18
Db	2698	ESTIQKILEKEQTPPNLASEPGRLLALKTGDEHILSTMHHAISDGHVSVDIFQOEI	2757	Db	3778	SYRYAAVLHVNDLAKPAHKVSPGAWVDFAAATKMDRDALIRLRGTKISHIAIANIPNSK	3837
QY	12	-----	11	QY	19	-----	18
Db	2758	QOFYSAILRGHDPLAQIAPLSIQYRDFATWQROIQFQVAEHRRLAYMTKQLADNKPABLL	2817	Db	3838	TIVERTICESVYDLGGDAKSDNRVSWLSAARSNAVKVASLSAIDLVDIAOAGFRVEIS	3897
QY	12	-----	11	QY	19	-----	18
Db	2818	TDFKRPMLSGRAGEIPVYVDGLIYEKLODFCRIRQVTAFTVLLAAFRAAHYRMTGTEDA	2877	Db	3898	CARQWSONGALDAVPHHLGSPSPOSSHVLIDFLTDHQRPEEALTNHPLHRAQSRRRVEROI	3957
QY	12	-----	11	QY	19	-----	18
Db	2878	TIGTPIANRRPELEGLIGFFVTCMRITVDVEDSFETLVHQRVRETTLAHAHANQDVPFE	2937	Db	3958	RERLOTLLPAYMIPAQIMVLDKPLPLNANGKVRKOLTQRAQTVPKAKQVSAPVAPRTEIE	4017
QY	12	-----	11	QY	19	-----	18
Db	2938	QIVSNILPGSSDTSRNPVLQMFALHSQNLGKVRLEGIEEBEIIISIAETTRFDIEFHLYQ	2997	Db	4018	RVLQCFSDVGLVDIGIMENFPDLGGHSLMATKLAARISRRLETHSVSKEIFDPHPRVCDL	4077
QY	12	-----	11	QY	19	-----	18
Db	2998	EAERLNGSIVVAADLFVPETIQSVITIFQILQKGLGEPDMPVSMALDGGLESRLSTGL	3057	Db	4078	VLIVQOGSAPHDPVSTKYTGVPVQSPAQGRLWFLDQLNFGATWYLMPLAVLRGAMNVH	4137
QY	12	-----	18	QY	19	-----	18
Db	3058	LHQOTDYPCDASVVOIFKQVAVNPVDIAVRDESTRLSYADLRKSDQVACWLSRRGIA	3117	Db	4138	ALTAALLALERHRELLRTTFYEQNGVMQKVPVVTETURIIDLSNGDGYLPTLKKEQT	4197
QY	19	-----	18	QY	19	-----	18
Db	3118	PETFVAILAPRSCETIVAILGVLKANLAYPLDVNPVAPASLEAILSEVSGSMLVLVGAET	3177	Db	4198	APHLETEPGRVALLRGLPGDYILSVMMHIIISDGSVDVLFQELGQFYSTAVKGHDPL	4257
QY	19	-----	18	QY	19	-----	18
Db	3178	PIEGMAEATIRITEILADAKTDDINGLAASQPTAASLAYVFTSGSTGRPKGVMEHR	3237	Db	4258	SQTTPLPIHYRDFALWQKPTQSEHERQLQYWVEQLVDSAPAELLTDLPRPSILSQAG	4317
QY	19	-----	18	QY	19	-----	18
Db	3238	GIVRLTKQTNITSKLPESFHMAHISNLAFDASVWEVFTLLNGGTLVCIDIFTLLLESTAL	3297	Db	4318	EMSVTIEGALYKNLEEFRCVRHVTSFVLLAALRAAHYRLTGSSEDATIGTPIANRRPEL	4377
QY	19	-----	18	QY	19	-----	18
Db	3298	EKVFFDQRVNVALLPPALLKQCLDNSPALVKTLSVLYIGDRDASDAKARGLVQTAQF	3357	Db	4378	EOIIGFFVNTQICRITVNEDETFESLVQOVRSTATAAFAHQDVPFEKIVSTLLPGSRDAS	4437
QY	19	-----	18	QY	19	-----	18
Db	3358	NAYGPTNTVMSTIYPIAEDPFIINGVPIGHAVSNSGAFVMDQNOQITPPGANGELIVTGD	3417	Db	4438	RNPLVQLMFAVHSQKNLGEKLENNAHSEVVPTEITTRFDLEFHLFOQDDKLEGSILYSTD	4497
QY	19	-----	18	QY	19	-----	18
Db	3418	GLARGYTTSSLNTGRFINVDIDGEQVRYRTGDRVRYRPRKDLQIEFFGGRIDHGVKIRGHR	3477	Db	4498	LFEAVSVQSLLSVFQEILRRGLNGPDVPISTLPLQDGIVDLQROGLLOVQKTEYPRDSSV	4557
QY	19	-----	18	QY	19	-----	18
Db	3478	IEPAEVEYALLSHDLVTDAAVTVHSQENQDLEMVGFVAARVADVREDESSNOVQEWQTHF	3537	Db	4558	VDVPHEQVSNPDSIALIHGSEKLSYAQLDRESDRVARWLHRHSFSDTLIAVLAPRSC	4617
QY	19	-----	18	QY	19	-----	18
Db	3538	DSIYADITTIQOQSLGRDFMSWTSNMYDGLIKKSQMQEWLDDTMRSLDSQPPGHVLEV	3597	Db	4618	TIIFLGLKANLAYPLDVKAAPAARIDAIVSSLPGKNKILLGANVTPPKLOEAAIDFVP	4677
QY	19	-----	18	QY	19	-----	18
Db	3598	GTGTGMVFNLRGEGGLQSVGLEPSPATAFVNKAAKSFPGLEDRIRVEVGTATDIORL	3657	Db	4678	IRDTFTTLTDGTLQDGTIERPSAQSLAYAMFTSGSTRPKGVVMVQHRNIVRLVNSNVV	4737
QY	19	-----	18	QY	19	-----	23
Db	3658	GDDLHAGLVVNSVAQYFSPDYLAQLVRDLTKVPGVERIFFGDMRSHAINRDFLVARAV	3717	Db	4738	AKOPAAARIAHISNLAFDASSWEIYAPLLNGAIVCADYFTTIDPOALQETFOEHEIRGA	4797
QY	19	-----	18	QY	24	-----	23
Db	3718	HALGDKATKAEIOREVVRMESEDELLVDPAFFTSLTTOVENIKHVIEILLPKMRATNELS	3777	Db	4798	MLPPSLLKQCLVQAPDMISRLDILFAAGDRFSSVDALQAOQLVGVGVFNAYGPTENTILS	4857
QY	19	-----	18	QY	24	-----	23

Db	4858	TIYNVAENDSFVNGVPIGSAVNSGAYIMDKXQQLVPAGVMGELVVTGGGLARGYMDPKL	4917
Qy	24	-----	23
Db	4918	DADRFIQLTVNGSEQVRAVRTGDRVYRPFKQFQIEFFGFRMDQOIKIRGHRIBPAEVEQAF	4977
Qy	24	-----	23
Db	4978	LNDGFVEDVAIVIRTPENOPEMVFVTAKGDSNAREEBEATTOIEGWEAHFEGGAYANIE	5037
Qy	24	-----EDT-----	26
Db	5038	EIESEALGYDFMGWTSWYDTEIDKDEMREWLNDTMRSLDDGPKAGRVLEVGTGTGMIMF	5097
Qy	27	-----BRYV-----	30
Db	5098	NLGRSGLERYIGLEPAPSAAEFVNNAKSFPGLAGRAEVHVGTAAVGTGLQGLTSDMAV	5157
Qy	31	-----	30
Db	5158	INSAQYFPTPEYLAETIKSLVQVPGMKRIYLGDMRSMWNRDFAAARAAYSLADNASKD	5217
Qy	31	-----	30
Db	5218	RVRQXMELEKEEELLVDPAPFTALASQLQDRIQHVEILPKMKATNELSSVRYAAVLH	5277
Qy	31	-----	30
Db	5278	ISDEPLPIYKIDPEAMINPEGSRLTREALAQLKENENAEVAISNIPYSKTVVERHIVR	5337
Qy	31	-----	30
Db	5338	SLQOEDANAPESMGSDMISAVTRAQOCHTSLASDLDIAEDAGFRVEVSWARQSOH	5397
Qy	31	-----LTN-----	33
Db	5398	GALDAVPHHLKPATEDSRVLIKPTDHOGRPLKSLTNQPLLPQASRRAEILLIREGLQTEL	5457
Qy	34	-----	33
Db	5458	PPYMIQSQITLIDRMLNANGKVDRELARRAKITOKSKPEVIDVPPRNSVEATVCKGFT	5517
Qy	34	-----	33
Db	5518	DVLGVEVGITDNFNLGSHSLMATKLAARLGRQLNTRISVRDVPDQVPVADLAAVIQNS	5577
Qy	34	-----LNIGA-----	38
Db	5578	APHEPIKPADYTGVPQSPAGRLWFLDQLNVGATWYLMPLGIRLHGLSRVDALATAISA	5637
Qy	39	-----	38
Db	5638	LEQRHEPLRTTFHEEDGVGVQVQDHRPKDLRIIDLSTOPKDAYLAVLKHEQTLFDLAT	5697
Qy	39	-----	38
Db	5698	EPGMRVALIRLGEHEHLSIVMHIIISDGWSVEVLFDEMHRFYSSALRQQDPMEQILPLP	5757
Qy	39	-----	38
Db	5758	IQRDFAAWQKTEEQVAEHQROLDYWEHLADSTPAELLTDLPRPSILSGRANELPLTIE	5817
Qy	39	-----	38
Db	5818	GRLHDKLAFRCRVHQATPFPVILLAAARAHYRLTGAEDATLGTPIANRNRELENMIGFF	5877
Qy	39	-----	38
Db	5878	VNTQCMRIAIBENDNFESLVRVRSTATSAFANQDVPFESIVSSLLPGSRDASRNPLVQV	5937
Qy	39	-----	38
Db	5938	ILAVHSQODLGKLTLEGLRDEAVDSAISTRDFVEFHLFEHADRLSGSVLYAKELFKLRTI	5997
Qy	39	-----	38
Db	5998	ESVSVFLETLRALDOPLTPLAVPLTDGVEIASKGLLDVPRDTPDYPRDANIVEVFOQH	6057
Qy	39	-----	38
Db	6058	VRATPDAIVKDATSILTYAQLDOQSDRLAIWLSRRHMPETLVGLVLAAPRSCETIIAMPG	6117
Qy	39	-----	38
Db	6118	IMKANLAYLPDINSIPAARLSILSAVDGNKLVLLGSGVTAPEQENPEVEAVGIOELAG	6177
Qy	39	-----	38
Db	6178	TGLDKTQGSNARPSATSLAYVIFTSGTGPKGVMEHRSVTRLAKPSNVISKLPQGARV	6237
Qy	39	-----	38
Db	6238	AHLANIAFDASIEIATTLNGATLVCLDYHTVLDRCRTLKEVFERESITVVVTLMPALLQ	6297
Qy	39	-----	38
Db	6298	CVAEIPETLAHLDLLYTCGDRVGGHDAMRARSILVKGHFGSGYGTENTVISTIIYEVDAD	6357
Qy	39	-----	38
Db	6358	MFVNGVPICKTVSNSGAYVMDRNOQLVPSGVVVGELVVTGDLARGYTDPSLNKNRFIVIT	6417
Qy	39	-----	38
Db	6418	VNGESIRAYRTGDRVYRPHDLOIEFFGRMDQOVKIRGRIEPEGEVESALLSHNSVQDAA	6477
Qy	39	-----	38
Db	6478	VVICAPADODSGAEMVAFVAAARNTEDETOEEAVDQVQGWETHFETAAYSEVKDIROSE	6537
Qy	39	-----	38
Db	6538	VGNDPMGWTSMYDGEIDKTDHMLNDTMRMILDAREPHGVLEIGTGTGMVFNLAACP	6597
Qy	39	-----	38
Db	6598	GLQGYGFEPESKAAQFVNDAAQSPALKDGRSIVHVGTTATDINKAGPIOPRLVVINSVA	6657
Qy	39	-----	38
Db	6658	QYFPTPEYLFVVVEALVQIPSVVERIVFGDMRTNAINRDFVASRALHTLGEKANKRLVRQM	6717
Qy	39	-----ELLRDPS-----	45
Db	6718	IYELEANEHEELTDPAPFTSLRGLGEKIKHVEILPKTMKATNELSKRYAAVLHVGRSR	6777
Qy	46	-----	45
Db	6778	EQSTIHQVSPNAWIDFAADGLDROTILNLLKEHKOAGTVAIGNIPYSKTIVERFVNKSLS	6837
Qy	46	-----	45
Db	6838	EDDMEEQNSLDGSAWAAVRAAQAQSCPSLDAMDVKEIAQEGYQVEVSWARQWSONGAL	6897
Qy	46	-----LGAQ-----	49
Db	6898	DAIFHHFEPPEKAGARTLIEFPTDYEGRVNVTLTNRPLNSIOSRRLLGTQIREKLOTLPPY	6957
Qy	50	-----	49
Db	6958	MIPSRIMVLDQMPVNNNGKIDRKELVRRRAIVAPKPSAATRAVAPRNEIEAILRDEFEDVL	7017
Qy	50	-----	49
Db	7018	GTEVSVLDNFFDLGGHSLMATKLAARVSRRLDAHISIKDVFQDQVLADLAASIQRESAPH	7077

Qy	50	----- 49	Qy	74	----- 73
Db	7078	EPIPORPYTGAQSFQAGRLWFLDQNLGATWYLMPLAIRGOLRVAALSAALFALER 7137	Db	8158	LWFLQNLGASWYLMFPFAIRMRGPIQTKALAVNALVHRHEALRTTFEDHGVQVI 8217
Qy	50	----- 49	Qy	74	----- 73
Db	7138	RHETIRTTFEESDGVGVQVGEARNSDLRVHDSVTGDDGGEYLEVLRRQTVPFDLSSPEG 7197	Db	8218	QPKSSODLRRIIDLSDAVDDTAYLAALKREOTTAFDLTSEPGWRVSLRLRGDDDDYILSVM 8277
Qy	50	FRVHLVK----- 56	Qy	74	----- 73
Db	7198	WRVCLVKTGEEDHVLISVMHHIYDGSVDILRGELGQFYSAALRGQDFLLHANPLPIQY 7257	Db	8278	HHIISDGTVDVLRQELGQFYSAAIRGOEPLSOAKSLPIOVDFAVWQROENIQEAKO 8337
Qy	57	----- 56	Qy	74	----- 73
Db	7258	RFAAWOREAKQVEEHQRLGYW\$KOLVDSTPAELLTDLPRPSILSGRAGSVDTIEGSV 7317	Db	8338	LKYWSQOLADSTPCEFLTDLPRPSILSGEADAVPMVIDGTQYQLLDFCRTHOVTSFVL 8397
Qy	57	----- 61	Qy	74	----- 73
Db	7318	YGALQSFCTRSTRTTFVLLTVFRIAHFRLTAVDATIGTPIANRRPELETLVGCFVNT 7377	Db	8398	LAAFRTAHYRLTGLDATVGTPTIANRRPELEGLIGFVNTQCHRMATSETETESLVQQ 8457
Qy	62	----- 61	Qy	74	----- 73
Db	7378	QCMRISIADDNFEGLRVORNATAAVANQDVPFERIVSALVPGSRNTRSRLVQLMFA 7437	Db	8458	VLTTTEAFANQDVPFEQIVSTLLPGSRDTSRNPLVQVMFALOSQODLGRIGLEMTDEA 8517
Qy	62	----- 61	Qy	74	----- 73
Db	7438	VOSVEDYQVRLEGLESVMPGEASTRFMEPHLVPGDKLTGSLVYSSDLPEQCTIQNF 7497	Db	8518	LETPLSTRLDLVHLFQEVGKLSGLLYSTDLFEVETIRGIVDVFLERLRRGLEBQPKORL 8577
Qy	62	----- 61	Qy	74	----- 73
Db	7498	VDIFOECLRSVLDQPLTPISVLFPFSNAISNLESLLDLEMTSDYPRDRTVVDLRFQAAI 7557	Db	8578	MAMPITDITKLRDOGLLTAKPAYPRESSVIDLPROOVAAPDAIAVWDSSSLTLYADL 8637
Qy	62	----- 61	Qy	74	----- 73
Db	7558	CPDSIAVKDSSSOLTVAOLDEQSDRVAAMLHERHMPAESLVGLSPRSCTIIAYFGIMK 7617	Db	8638	DQSNKLAHMLCQRMAPETLVAVFAPRSCLTIIVAPLGVLKANLAYLPDYNAPARIEA 8697
Qy	62	----- 61	Qy	74	----- 73
Db	7618	ANLAYPLDVYAPDARLAAILDTVEGERLLLLGAGVQPQIQIPLRSTAYIAEALSHATT 7677	Db	8698	ILSAVPGHKLVLQAHGPELGLTWADTELVDALASSSSGDHEQIHASGPTATSAYV 8757
Qy	62	----- 65	Qy	74	----- 73
Db	7678	VDVTSIQPSATSALAVIFTSSTGPKGVMIEHRGIVRLVDTNVNVPFSGSALPVSH 7737	Db	8758	MFTSGSTGPKGVMIDHRSIIRLVNSDDVATLTPVRMANVSNLAFDISVQEIYATLLN 8817
Qy	66	----- 65	Qy	74	----- 73
Db	7738	FSNLWDAATWEIYTAVLNGTVVICIDRTMLDIAALNSTFRKENVRAAFPTPAFLKQCL 7797	Db	8818	GGTLVCLDYLTLDSKILYNNVFEAQVNAAMFTPVLLKQCLGNMPAIIISRLSVLFNVGDR 8877
Qy	66	----- 73	Qy	74	----- 73
Db	7798	AETPELVANLEILHTAGDRLDPGDANLAGTKAGGIFNVLGHTEHTAYSTFFPVVGEETF 7857	Db	8878	LDAHDAVAASGLIQDADVYNAYGPTENGHQMSTMYKVDVNEPFVNGVPIGRSITNSGAYVMD 8937
Qy	74	----- 73	Qy	74	----- 73
Db	7858	VNGVPVGRGISNSHAYIIDRHQKLPAGVMGELLITGDGVARGYTDNALNKDRFVVYDIN 7917	Db	8938	GNQQLVSPGVNGEIVVTGDLARGYTDSALDEDRFVHTIDGEEENIKAYRTCDRVRYRPK 8997
Qy	74	----- 73	Qy	74	----- 73
Db	7918	GKSTWYRTGDKARYRPRDQGLEFFGRMDQMWKIRGVRIEPEGEVELTLLDHKSVALAATVV 7977	Db	8998	DPEIEFFGRMDQVKIRGRIEPAEVEHALLCHDLVHDAAVVLRKPAQNEPIAFITSO 9057
Qy	74	----- 73	Qy	74	----- 73
Db	7978	VRRPPNGDPENIAFITIDAEDDVQTHKAIYKHLOGILPAYMIPSHLVILDQMPVTONGKV 8037	Db	9058	EDETIEOHESNKQVQGWGEHFDVSRYADIKOLDTSTFGHDFLGWTSMDYGVDDIPVNEKME 9117
Qy	74	----- 73	Qy	74	----- 78
Db	8038	DRKDLALRAQTVQKRRTAARVPRDEVEAVLCEEYSNLLVEVEGVTGDFDLGCHSLLA 8097	Db	9118	WLDETTASLLDNRRPPCHILEIGAGTGMILSNLKGKVDGLQKYVGLDPAPSAAFVNEAVKS 9177
Qy	74	----- 73	Qy	79	----- 78
Db	8098	TKLAARLSQLNTRSVKVDOPILADLADIIRRGSHRHRDPIPATPYTGPVEQSFQAGR 8157	Db	9178	LPSLACKARVLGVTALDIGSLDKNEIQPELVVINSVAQVFPPTSEYLIKVKVKAWEVPSVK 9237
			Qy	79	----- 78

Db 9238 RVFFGDIRSOALNRDFLAARAVPALGDNASKEQIREKIAELESESEBELLVDPAFFVLSRS 9297
QY 79 ----- 78
Db 9298 OLPNIKHVEVLPKMKATNELSSRYAAVLHISHNEEQLLIQIDPTAWVDFAATQKDS 9357
QY 79 -----SVC 81
Db 9358 QGLRNLLQGRDDVMIAVGNIPYSKTIIVERHIMNSLDQDHVNSLDGTSWISDARSAAAI C 9417
QY 82 ----- 81
Db 9418 TSFDPALQLAKEEGFVELSWARQSQNGALDAVFHRLATDANCERSVLVHPTDQ 9477
QY 82 ----- 81
Db 9478 GRQLRTLNRPLQAOSSRIESQVFEALQALPAYMIPRSIIIVLPQMPNTANGKVRKQL 9537
QY 82 ----- 81
Db 9538 ARRAQVAKRKAVSARVAPRNDTEIULCEEYADILGTEVIGITONFPDMGHSMLATKLAA 9597
QY 82 ----- 81
Db 9598 RLSRLDTRVTYKVEFDKPVLAADLAASIEOGSTPLPIASSVYSGPVEQSYAOGRLWFLD 9657
QY 82 ----- 81
Db 9658 QFNLNATWYHMSLAMRLGLPLNMDALDVALRALEQRHETLRTTTEAOKDIGVQVHEAGM 9717
QY 82 ----- 81
Db 9718 KRLKVLDSLDPKEHEHMAVLENEQMRPFTLASEPGMKGHARLGPTGYILSLVMHMFSD 9777
QY 82 GWS----- 84
Db 9778 GWSVDILRQELGOFYSNALGRDPLSQVKPLPIQYRDFAAWQKAAQVAEHERQLAYWEN 9837
QY 85 ----- 84
Db 9838 QLADSTPGLLTDPRPOFLSGKAGVIPVTIEGPVYKLLKFSKEROVTLFVLLTAFRA 9897
QY 85 ----- 84
Db 9898 THFRLTGAEDATIGTPIANRNRPELEHIIGFFVNTQCMRLLLDTGSTFESLVQHVRSVAT 9957
QY 85 ----- 84
Db 9958 DAYSNOPIPERIVSALLPGSRDASRPLIQLMPALHSQPOLGNITLEGLEHERLPTSVA 10017
QY 85 -----QTIN----- 88
Db 10018 TRFDMEPHLFOEPNKLGSILFADELFOPETINSVTVFOEILRRGLDQPVSI STMPLT 10077
QY 89 ----- 88
Db 10078 DGLIDLEKLGLEIESSNFRDYSVDVFRQOVAANPNAPAVVDSETSMSYTSLDQKSEQ 10137
QY 89 -----PE----- 90
Db 10138 IAAWLHAOGLRPESLICVMAPRSPETIVSLFGILKAGYAVILPDVNSPAARIQPIELSEVE 10197
QY 91 ----- 90
Db 10198 GKRLVLLSGIDMPQSDRMDVETARIQDILTNTKVERS DMPRSATSLAYVFTSGTG 10257
QY 91 ----- 90
Db 10258 RPKGVMIEHRNLRVQKSNVTSQLPDLRMAHISNLAFAASIWEIFTAILNGALICID 10317
QY 91 ----- 90

Db 10318 YFTLLDQALRTTTFEKARVNATLFPALLKBLCLNHAPTLFEDLKVLYYIGGBRLDATDAK 10377
QY 91 ----- 90
Db 10378 IOALVKGTVYNYAGPTENTVMSTIYRLTDGESYANGVPIGNAVSSGAYIMDOKQLVPP 10437
QY 91 ----- 90
Db 10438 GVMGELVVGDLGARGYTNSTLNADRFVDIVINDQKARAYRTGDRTRYPKDGSIFFGR 10497
QY 91 ----- 90
Db 10498 MDQOVKIRGHRVEPAEVEQAMLGNKAIHDAAVVQAVDQGETEMIGFVSMASDRSEGBE 10557
QY 91 -----DDT----- 93
Db 10558 EITNQVQEWEDHFESTAYAGIEAIDQATLGRDFTSWTSMYNGNLIDKAEMEELDDTMS 10617
QY 94 ----- 93
Db 10618 LLDKEDARPCAEIGTGTGMVLFNLPKNDGLESYVVGIEPSRSAALFYDKAAODFPGLQKKT 10677
QY 94 ----- 93
Db 10678 QILVGTAEIDIKLVDFHPDVVIVNSVAQYFSPRSYLVQIASELIHMTSVKTIFFGDMRSW 10737
QY 94 -----DP----- 95
Db 10738 ATNRDPLVSALYTLGDKATKQIQREVARLENEDELLVDPAFTSLTSQWPGKYKHYE 10797
QY 96 ----- 95
Db 10798 ILPKMRTSNELSSYAAVLHI CRDGEGRNRYGRRVHSVEENAWIDFASSGMDRHALVQ 10857
QY 96 ----- 95
Db 10858 MLDERRDAKTAIGNIPHSNTINERHFTSLDTEGEGIAODSLDGSAMOSATKAMARCP 10917
QY 96 ----- 95
Db 10918 CLSVTELVEIGQAAGFRVEVSWARQSHGALDVVPHHLEDDRVGRVLINFPDTERLPP 10977
QY 96 ----- 95
Db 10978 STGLTSRPLQIRNRRFESQIREQLTLLPYMVPSRIVVLERMPNLANSKVDRELARK 11037
QY 96 -----GH----- 97
Db 11038 ARTLOTIKPSATRVAPRNDIEAVLCDEFOAVLGVTVGMNDNFELGGHSLMATKLAARLS 11097
QY 98 -----ADLV----- 101
Db 11098 RRLDTRVSVKIDFNQPILODLADVOTGSAPHEAIPSTPYSGPVEQSFQGRMLFLDQLN 11157
QY 102 ----- 101
Db 11158 LNASVHYMPLASRLRGLRIEALQSALATIEARHESLRTTFEBQDGVVPQIVAAARNKOL 11217
QY 102 ----- 101
Db 11218 RIIDVSGTEDAYLAALKQEQDAADFDTAPBGRVALLRLGPDHVLUSIVMHHIISDQWSV 11277
QY 102 -----LY----- 103
Db 11278 DILRQELGQLYSNASSQAPLPIQYRDFAIWQKDSQIAEHOKQLNWKYKQLVNSKPAEL 11337
QY 104 ----- 103
Db 11338 LADFTRPKALSGDADVPIEIDDOVYQNLRSCFRARHVTFSVALLAAFRAAHYRLTGAED 11397
QY 104 ----- 103
Db 11398 ATIGSPIANRNRPEGLIGCFVNTQCLRIPVKSEDTFTDLVKQARETATEADQNQDVPF 11457

```
Qy 104 -----ITRDELE--- 111
Db 11458 ERIVSSMVASSRDTSRNPLVQVFAVHSQHDGNIRLEGVEGKPVSMAASTRFDEAHLP 11517
Qy 112 ----- 111
Db 11518 EDQMLGNVVFVKDLFESETIRSVAVFOETLRRGLANPHANLATPLTDGLPSLRSLC 11577
Qy 112 ----PD----- 113
Db 11578 LOVNQDYPDRDASVIDVREOVASIPKSIADIVASSOLTYTELDERSSQATWLRQVTV 11637
Qy 114 -----GNQV----- 118
Db 11638 PEELVGLAPRSCETIIAFLGIKANLAYLPDYNAPAGRIETILSLPCNRLILGSDT 11697
Qy 119 -----R 119
Db 11698 QAVKLHANSVRFRISDALVESGSPTEELSTRPTAQSLAYVMFTSGTGVPKGMVVEHR 11757
Qy 120 GVT----- 122
Db 11758 GITRLVKNVNAKOPAAAAIAHLSNIAPDASSWEIYAPLLNGTVCIDYYTTIDIKAL 11817
Qy 123 ----- 122
Db 11818 EAVFKQHIRGAMLPPALLKQCLVSAPTMISSLEILFAAGDRSSQDAILARRAVSGVY 11877
Qy 123 ----- 122
Db 11878 NAYGPTENTVLSTIHNIGENAEFSGVPIGNAVSNSGAFVMDQONQQLVAGVIGELVWTG 11937
Qy 123 ----- 122
Db 11938 DGLARGYDTSKLRVDRFIYITLDGNRVAYRTGDRVHRPKQOIEFFQGMQOIKIRGH 11997
Qy 123 ----- 122
Db 11998 RIEPAEQALARDPAISDAVITQUTDEEPELVAFSLKGNANGTNGVSGDQEKID 12057
Qy 123 ----- 122
Db 12058 GDEQHALLMENKIRHNLQALLPTYMIPRIIHVDQLPVNANGKIDRNEIAVRAQAATPRTS 12117
Qy 123 ----- 122
Db 12118 SVSTYVAPRNDIETIICKEFADILSVRVGIDTNFDFLGGHSLIATKLAARLSRRLDTRVS 12177
Qy 123 ----- 122
Db 12178 VRDVFDPVVGOLAASIOGGSTPHEAIPALSHSGPVQOSFAQGRFLWFLDRFNLAAWYIM 12237
Qy 123 ----- 122
Db 12238 PFGVRLGRLVDALOTALAEERHELLRTTFFEQDQVGMQIVHSPRMRDICVVDISGA 12297
Qy 123 ----- 122
Db 12298 NEDLAKKEEQQAPFNLSTEVAVRALFKAGENHHLSIVMHHIISDQMSVDIFQQLAQ 12357
Qy 123 ----- 122
Db 12358 FYSVAVRGHDPLSQVKPLPIHYRDPFAVWORQKQVAVHESQLQYWIEQLADSTPAEILSD 12417
Qy 123 ----- 122
Db 12418 FNRPEVLSOEAQTVPIVIEDEVYKLSLFCRNHQVTSFVLLAARVAHYRLTGAEDATI 12477
Qy 123 ----- 122
Db 12478 GTPIANRNRPELEDLIGFFVNTQCMRIALEEHDNFLSVVRRVRSSTAASAFENQDVPERL 12537
```

```
Qy 123 ----- 122
Db 12538 VSALLPGSRDASRNPLVQLMFVHSORNLGKLEGELEGEPTPYTATTTRDFVEFHLFEOD 12597
Qy 123 ----- 122
Db 12598 KGLAGNVVFAADLFEAATIRSVVEFHEILRRGLDQPDIAISTMPLVDGLAALNSRNLP 12657
Qy 123 ----- 122
Db 12658 VEDIEPDFAEASVVDVFQTVVANPDALAVTDTSTKLYAELDOQSDHVAWLSKQKLP 12717
Qy 123 ----- 122
Db 12718 ABSIVVVLAPRSSETIVACIGILKANLAYLPWDSNVPEARQAIISEIPGEKFFVLLGAGV 12777
Qy 123 ----- 122
Db 12778 PIPDNKTADVVRMFISDIVASKTDSYSGPTRPSASSLAYVIFTSGSTGRPKGMVVEHRG 12837
Qy 123 ----- 122
Db 12838 VISLVKQNASRIPOSLRMAHVNLAFDASVWEIFTLLNGGTLFCISYFTVLDKALSAA 12897
Qy 123 ----- 122
Db 12898 FSDHRINITLLPALLKQCLADAPSVLSSLESLEYIGDRLDGADATKVKDLVKGKAYNAY 12957
Qy 123 ----- 122
Db 12958 GPTENSVMTIYTIETIETETANGVPIGTSLGPKSKAYIMDQOQLVPAGVMGELVAGDGL 13017
Qy 123 ----- 122
Db 13018 ARGYTDPSLNTGRFIHITIDCKQVAYRTGDRVRYRPRDYOIEFFGRLDQOIKIRGHRIE 13077
Qy 123 ----- 122
Db 13078 PAEVEQALLSDSSINDAVVSAQNKEGLEMVGYITTOAQSDVKEEASNKVQEHAFDS 13137
Qy 123 ----- 122
Db 13138 TAYANIGIDRDALGODFLSWTSMYDGLIPREEMOEWLNDTMRSLLDNOPPGKVLBIGT 13197
Qy 123 ----- 122
Db 13198 GTGMVFNLCVREGLOSAGLEPSRSVTAWNKAIEFFPSLAGSARVHVGTAEIDISSIDG 13257
Qy 123 ----- 122
Db 13258 LRSLDVVINSVAQVFPFSREYLAELTANLIRLPGVKRIFFGDMRTYATNKDPLVARAVHTL 13317
Qy 123 ----- 122
Db 13318 GSNASKAMVROQVAKUEDDEEELLVDPAPFTSLSDQFPDBEIKHVEILPKRMAATNELSSY 13377
Qy 123 ----- 122
Db 13378 RYAAVHVGHCHOMPGEDEDEKQAVKDINPKAWVDFAGTMDRQALLQLLQDRQRGDDVV 13437
Qy 123 ----- 122
Db 13438 AVSNIPYSKTIMERHLSQSLDDDEDDGTSADVGTAMISRTQSRACEKPCALSVADLIEIGK 13497
Qy 123 ----- 122
Db 13498 IGFEVEASWARQHSORGGDLDAVPHRPFPPRHSCHVMFRPPTTEHKGRSSSLTNRPLHLLO 13557
Qy 123 ----- 122
Db 13558 SRRLEAKVRERLQSLLPYMPIPSRITLLDQMLPTNSGKYDRKKLARQARVIPSAASTLTD 13617
Qy 123 -----OLGG----- 126
```

Db 13618 FVAPRTEIEVLCFEFTDLGVGVGIDTNFFELGSHSLATKLSARLSRRLDAGITVKQV 13677
QY 127 ----- 126
Db 13678 FDQVPLADLAASILQSSRRHSIPSLPYEGPVEQSFQAQRLWFLDQFNIDALWYLIIPFAL 13737
QY 127 ----- 126
Db 13738 RMGPLOVDALAAALVALEERHESLRITTFEERDGVGIQVQVPLRTTKDIRIIDVSGMRDD 13797
QY 127 ----- 132
Db 13798 DAYLEPLQEQQTDFDLASEPFGWRVALLKLGKDDHILSVMMHHIISDGMSTEVLORELOQ 13857
QY 133 ----- 132
Db 13858 FYLAASKGKAPLSQVAPLPIQYRDFAVWQREQVAESQRLDYWKQOLADSSPAELLAD 13917
QY 133 ----- 132
Db 13918 YTRPNVLSGEAGSVSFVINDSVYKSLVSFCRSQVTTFTTLAFAAAHYRMTGSDDATI 13977
QY 133 ----- 136
Db 13978 GTPIANRNPENLIGCFVNTQCMRITIGDDEFSLVQOVRSITTATAFENODVPFERI 14037
QY 137 ----- 144
Db 14038 VSTLSAGSRDTSRNPLVOLLFVAVHSQQGLGRIQLDGVVDEPVLSTVSTRDLSEFAHQEA 14097
QY 145 ----- 144
Db 14098 DRLNGVMFATDLFQPETIQGVAVVEVLQGLEQPOSPIATMPLAEGIAQLRDAGALQ 14157
QY 145 ----- 144
Db 14158 MPKSDYPRNASLVDFVQOQNASPSTVAVTSDTSKLTVAELDRLSQOAAASYLRROQLPAE 14217
QY 145 ----- 144
Db 14218 TWAVLAPRSCETIIAFLAILKANLAYMPLDVNTFSPARMEAIISVPCRRLLILVSGVRH 14277
QY 145 ----- 144
Db 14278 ADINVPNAKTMILSDVTGTDAIGTPEPLVVRPSATSLAYVIFTSGTGKPKGVMEHRA 14337
QY 145 ----- 144
Db 14338 IMRLVKDSNVVTHMPATRMARHTNIAFDVSLFEMCATLLNGGTLVCIDYLTLLDSTMRL 14397
QY 145 ----- 144
Db 14398 ETFEREQVRAAIFPPALLRQCLVNMPPDAIGMLEAVYVAGDFHRSRDARATQALAGPRVYN 14457
QY 145 ----- 144
Db 14458 AYGPTENAILSTIYNIDKHPYVNGVPICSAVNSGAYVMDRNQQLPPGVNGELVWTGE 14517
QY 145 ----- 144
Db 14518 GVARGYTDASLTDREFTVITDQQRQAVRTGDRVRYRPKFQIEFFGRLDQQAQIRGHR 14577
QY 145 ----- 144
Db 14578 VELGEVEHALLSENSVTDAAVLRITMEEDPOLVAFVTTDHEYRSGSNEEDPYATQAA 14637
QY 145 ----- 144
Db 14638 GDMRKLRLSLLPYVMVPSRVTILROMPLNANGVKVDRKDLARRAQMTPTASSSGPVHVAPR 14697
QY 145 -----GVT 147
|:|

Db 14698 NETEAALCDEFETILGVKVGIT 14719
RESULT 7
Q93NW6 PRELIMINARY; PRT; 10917 AA.
ID Q93NW6
AC Q93NW6
DT 01-DEC-2001 (Tremblurel. 19, Created)
DT 01-DEC-2001 (Tremblurel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblurel. 21, Last annotation update)
DE AMPHC.
GN AMPHC.
OS Streptomyces nodosus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=40318;
RN (1)
RP SEQUENCE FROM N.A.
RA Caffrey P., Lynch S.V., Flood E.M., Finnian S.M., O'Liynk M.;
RT "The amphotericin biosynthetic gene cluster from Streptomyces
nodosus.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF357202; AAK73514.1;
DR InterPro; IPR001227; AC transferase.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR004410; Fabb.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppanthe_attach.
DR Pfam; PF00698; Acyl_transf_6.
DR Pfam; PF00107; adh_zinc; 1.
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF00550; PP-binding; 6.
DR TIGRfam; TIGR00128; fabb; 6.
DR PROSITE; PS00075; ACP DOMAIN; 6.
DR PROSITE; PS00509; B_KETOACYL SYNTHASE; UNKNOWN 1.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; UNKNOWN 6.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN 5.
SQ Phosphopantetheine.
SQ SEQUENCE 10917 AA; 1132905 MW; 15AC5956BBS810A1 CRC64;
Query Match 31.4%; Score 243; DB 2; Length 10917;
Best Local Similarity 1.1%; Pred. No. 21;
Matches 94; Conservative 21; Mismatches 33; Indels 8751; Gaps 27;
QY 1 AAG----- 3
Db 384 AAGVKNMAIQHGVLPRLSHAERPTSNVDWTAGAVELLDRAVDMPETGRARRAGVSFG 443
QY 4 ----- 3
Db 444 ISGTNAHLILEQAPTAPERPADPAARPALVPMTLSGRTAAALDAORERLAHSHPGCA 503
QY 4 ----- 3
Db 504 DPLDIGFALADGRATFEHRAVLACGTASAAATEIAHGVSVDGASAVLFGSGGSRORLGMG 563
QY 4 ----- 3
Db 564 RELHARFPVFAEFADEIAALLDRHTRPLREVVYVWGSDAELLNETGWTOPALFAVEVALYR 623
QY 4 -----GILHLE----- 9
Db 624 LVSSLGVTPTYVGHSIGEIAAAHVAAGVLSLEDACALVAARARLMQELPRGGAMLAVRAT 683
QY 10 ----- 9
Db 684 EEEVAPHLTERVSIASVNGPSSVVVAGDEDEVAIAAHFEAQDRKTTTLRLRVSHAFHSPLM 743
QY 10 ----- 9
Db 744 DPMLEEFRRVARGLTYHEPRIPVVSNTLGTADYADPADLCTADYVVRHVREAVRFADGVTTL 803

Qy	10	-----LLVAVGPD-----	17
Db	804	ACRGVTTLLEVGPDGVLSSAMAQESLPEGTTGVPLLRKDRPEESALTGLRVHVRGLTVR	863
Qy	18	-----	17
Db	864	WARLFDSAGAGRADLPTYFONQVMPAAGRARTSPAAAGLSADHPLLGAAVELADGEG	923
Qy	18	-----	17
Db	924	FLFTSRLSVSHPLADHGVQAAPCVPGTAFVELAIRAGDEAGCDRVEBELTLAAPLVLPE	983
Qy	18	-----	17
Db	984	RGAVQLQVRVGAPDPAGRRTLTGTSERVEDGFDLPMSQHATGVLTTCAGAPDAADFATVMP	1043
Qy	18	-----	17
Db	1044	PAGAEQVDLDGCRYERLAGLFDYGPAPRALRAAMRRGTEVFAEVALSENAGTDPADFGLIH	1103
Qy	18	-----	17
Db	1104	PALLDAAQHAAAYNDLGAISRGGLPFAMEGVHAAAGATTVRARITSVGEDTVSIAYDT	1163
Qy	18	-----VFQ-----	20
Db	1164	AGGTVLSVDLSVRAVPKDIPTCARAVHRDSLFOVDWTPLRQSQGLPETVALVGADPAG	1223
Qy	21	-----	20
Db	1224	LADTLRTGCVQTVHPDLAALAADPAPVPGTVLTALPGDQDTPAADAHTTAAAVLGLAQ	1283
Qy	21	-----AHQEDTERYVLTNL-----	34
Db	1284	HMLTEDRFADARLVLTTRDATAGDLAAAAGGLIRTARTENPGRFVLLDLATGTPAAPA	1343
Qy	35	-----	34
Db	1344	ALVAALAAAYDEPDLAVRGSDVLTARLARVPADAGADWNPDGTVLITGGTGGLGGVLAR	1403
Qy	35	-----	34
Db	1404	HLVTERGVRRLLASRRGPAADGAPALVDELTAALGASVDVAACDVTDREALAALLARVPA	1463
Qy	35	-----	34
Db	1464	AHPLSAVVHTAGVDDGVIGSLTAERLOTVLRPKADAAWHLHELTRDLDLDAFVLFSSVA	1523
Qy	35	-----	34
Db	1524	ATLSPGOGNYAAGNAFLDALAAHRAHGLPATSLAWGPWTQSVGWTGTLSDIDVERIAR	1583
Qy	35	-----	34
Db	1584	SCMPPLTVEQVALFDAALATGCPALLPVRLDLAVLRTQGEIAPLLRGLIRTTVRRAAAQ	1643
Qy	35	-----	34
Db	1644	VSVTADGLAQQLAGLADAAAREALLGLVRTQIAVLGHADAGQVEAARQFQDLGFDLSLA	1703
Qy	35	-----	34
Db	1704	VELRNLNSATGLRLPATWVFDPYTPNALADHLSDELGTGONIPLPTPAGALPATDDP	1763
Qy	35	-----	34
Db	1764	IVIVGMACRYPGCVTSPEDLWRLVSEGADATGFPFVNRGWDVNLVDPPDPRGHTYVRA	1823
Qy	35	-----	34
Db	1824	GGFLHDAGSFDADFFGMSPREALSTDSQORLLELSWEAVERAGIDPHSLRGSRTGVFAG	1883
Qy	35	-----	34
Db	1884	VYNDYGTILTGEYEAPRGNSAPSVASGRVSYTLGLEGPAVTVDTACSSSLVGMHLAA	1943
Qy	35	-----	34
Db	1944	QALRTGECTLALAGGVTVMTSTFTVDFSRQRLAADRSKAFAPAAADVAMSEGVGMLL	2003
Qy	35	-----	34
Db	2004	LERRSDAIRNGHEILAVLRGSVNDGASNGLTAPNGPSQQRVIRQALASGGLSPEDVDV	2063
Qy	35	-----NIG-----	37
Db	2064	VEAHGTGTTLGDPIEAQALIAIATYGRDRDPERPLMLGSKVKNIGHTQAAAGVAGVIKIMS	2123
Qy	38	-----	37
Db	2124	MRHGLLPOTLHVDAPSSHVDWAGAVELLTEOTEPENDRVRRRAVSVSGISGNTAHVII	2183
Qy	38	-----AELLRD-----	43
Db	2184	EQPAAGPAPAAPSPGAEELPAAPWPLSAKTPEALRDQARLLAHVEAHPETRPADIAVSL	2243
Qy	44	-----	43
Db	2244	LTSRSVPDHRAAVLGDRTEALRALAALAGTDDALLTGRTARGORTAFLFSGOGSORLG	2303
Qy	44	-----	43
Db	2304	MGRBLYGRVPVTEALDAVLAVLDGELEGSUREVMGEDADLLNETGTQTPALFAVEVAL	2363
Qy	44	-----	43
Db	2364	YRLVESGVKPDFVAGHSIGIEIAAAHTAGVLTADACRLVAARAALMOALPEGGMIAVE	2423
Qy	44	-----	43
Db	2424	ATEDEVLLPLTGDVSIAAVNGPSSVVVSGAEAAALAVAGHFTGCGRTSRLRVSHAFSP	2483
Qy	44	-----	43
Db	2484	LMDPMLDEFRSVAEGLSYSAPAI PWVSNLTGTLADPADLCSADYVVRHVRDAVPADCVG	2543
Qy	44	-----	43
Db	2544	TLGEHGVTTFVELGPDGVLSSMAQOQSVTEGAVTVPLRKDRSEBQALVTALCRLVVLGID	2603
Qy	44	-----	43
Db	2604	ADWTSALTVCGAARVDLPTVAFQHRWFMPAARAAHPDMRAAGLGSASHPLLLGAAVELPG	2663
Qy	44	-----	43
Db	2664	GEVLLTGLRSLQSHPLADHAVHGTVLLPGTALLDLAIRAGDEAGCGHLEDTLAAPT	2723
Qy	44	-----	43
Db	2724	LPQDGA VLLQVRVGAADDAGRRVTTVHARPEDATDGVWTLHATGTLSDTAATGALTDTV	2783
Qy	44	-----PSLGAQ-----	51
Db	2784	WPAGAQAALDTGECYQRFPAHGFAYGPLFQGLTAAMKAGEALYAEIALPESGHDDAAGFA	2843
Qy	52	VH-----	53
Db	2844	VHPALLDAAHPALLTDDGGLPFSWENSVLHSGATALRVLAPAGANALSVTAADPAG	2903
Qy	54	-----	53
Db	2904	RPVVSIARLIARPVVEADOLTGDAGHLRDALFRLEWTPVPLPDGTPTESALLGPDTDLLA	2963
Qy	54	-----	53

Db	2964	EALGEGAVRYATLADLAADAPVPGAVLVPLTAGSDTDDTPGAVHRTAAALTLLQEWLA	3023
Qy	54	-----	53
Db	3024	EDRFNGSRLVLRGAVATDDAPVTDLAAAALVGLVRSQAQENPGVGFMDLDADGASDS	3083
Qy	54	-----	53
Db	3084	ALTRALGLDEPOLALRGGRALAARLTRATADSTARVAMSSEGTVLVTGGTGGGLGLVARH	3143
Qy	54	-----	53
Db	3144	LVMERGVRRLLLTSSGLGAAGAGDLVAELEGAEVSVAACDVARDAVTALVAAPAE	3203
Qy	54	-----	53
Db	3204	HPLVAVHTAGVLDGVDLSLTERLATVLRPKVDGNWNLHEATRLDLDAFVVVFSSVAG	3263
Qy	54	-----	53
Db	3264	VFGAGQANYAANAFLDALTAAHRLGLPAVSLAWPMQOGAGMTAGLDERDIRRAAES	3323
Qy	54	-----	53
Db	3324	GMPLITVEOGLAFDAALTDEAALVTSRLDLAAFRGRPEIPALLRGLIRTPVRRAAATG	3383
Qy	54	-----	53
Db	3384	AADGADAGLVORLAGLDREORHEVLLDLVRTSAALVLGHAAGSDIDADRAFDLGFDSLT	3443
Qy	54	-----	53
Db	3444	AVELNRLRTATGURLSATMVFDHPTLSALADHLLTFLFGAEATLTPVTLPTDDPI	3503
Qy	54	-----	53
Db	3504	VIVGACRFPGDVTSPEDLWRLVSEGTDAISDFPANRGWVDLSLYHPDPHSGTSYTRSG	3563
Qy	54	-----	53
Db	3564	GFLHDAGDFDPDFGMSPREALATDSOORLLETSSWEAIERAGIDPLSLRDSRTGVFAGV	3623
Qy	54	-----	53
Db	3624	MYHDYSLGKKEFEGFQGSAGSVASGRVSYTFEGEPATVDTACSSSLVALHMAAQ	3683
Qy	54	-----	53
Db	3684	ALRSGECSLALAGGVTVNMSTPTFIEFSRQOGLSADGRSKAFESADGVGHSEGVGMILL	3743
Qy	54	-----	53
Db	3744	ERQSDAIRNGHEILAVVRGSAVNQDGASGLTAPNGPQORVIRQALASGGLSTGDVDAV	3803
Qy	54	-----LVKM-----	57
Db	3804	EAHGTGTTGLDPIEAQALIATYQONRTDEQPLWLSLKSNIIGHTQAAAAGVAGVKNWMSM	3863
Qy	58	-----VILT 61	
Db	3864	RHGVLPTLHVDAPSSHVDWTEGAVELLTEQTEWPETDVRRAAVSSFGISGTNAHVILE	3923
Qy	62	EPEGAP- :	67
Db	3924	QPVWAPATADVGAEPAPVPLILSGKSPALRDOAARLLDVTREHTALRPLDLGHSIA	3983
Qy	68	-----	67
Db	3984	TSRSAPDHRVAVLATGREDALRALTALADDEANSAVTCRTSRGRRALFSCGQSQRLCM	4043
Qy	68	-----	67
Db	4044	GRELYGRFPVFAEALDAVLADGELEGSUREVMGSEDAGLLNETGTQPALFAVEVALY	4103
Qy	68	-----	67
Db	4104	RLVESWGVPRPDFVAGHSIGEIAAAHAGVFSLEDAARLVAAGRLMOALPAGGAMVAQOA	4163
Qy	68	-----	67
Db	4164	TEDEVIPYLSDEVSI AALNGPASVVVSGAEGAVLEVAARFEAEGRKATMLRVSHAFHPL	4223
Qy	68	-----NITANLTSSL- :	77
Db	4224	MDMLDEPRAVETLSFAAPVIPVSNLTGSLATAEELCSPEYVVRHVREAVRFADGUSA	4283
Qy	78	-----	77
Db	4284	LEAQGVTTTFLGLPGDGVLSMAQESLTGDEAVTVALLRKORDEBETAALATALAALHTAGVR	4343
Qy	78	-----	77
Db	4344	LDWNAFFAGTGATRVLDLPTYAFQHTTYWPSGTHSADAAAAGVLTAAAGHPLLNGSVOLAEG	4403
Qy	78	-----	77
Db	4404	EGVLTGRLSLQSHPLADHAVHGTVLLPGTALLDLAPRAGDEVCCDRVEELTLAAPLVL	4463
Qy	78	-----	77
Db	4464	PERGAVQTVRGVADDTGRRRTTVVHSRPEDATDVSTQHTATGTLMPGTARTDGTGDTATV	4523
Qy	78	-----	77
Db	4524	WPPTGAELDTBGCYERFTALGFGYGPVFGQLKAAWRSGDVLVYAEVLOEAGAADDEAGF	4583
Qy	78	-----L 78	
Db	4584	GLHPALLDAAHALHASLVAHEGEGNGGLPFAWEGAVLHATGATALRVLTPTDEARQSVAI	4643
Qy	79	SVC- :	83
Db	4644	AVCDTAGQPVASIDRLLVRAVTDGDLNEAAGIARDALFALGWTPLPQDGDIPAAVALVGP	4703
Qy	84	-----	83
Db	4704	DTTGLAALTAAGAAVTAHQDLAALAAADGTDIPGTVVTVDTGTAAPVABAAHTVAAEALA	4763
Qy	84	-----	83
Db	4764	LAORWTAEDRFAGARLVFATCGAVEAGGSQVTDVAAAAMVGLVRSQAQSESPDTFVLVDRE	4823
Qy	84	-----	83
Db	4824	AADAAGDAABERGSLFLAALGSGEPQLARDDVVLAGRLARFDAAGANTLPAERAWHLDS	4883
Qy	84	-----	83
Db	4884	TPGSINGLALTPHPEALEPLTGHQARI EVRAAGLNFRDVLKALGNYPGDAGRYGHEAGV	4943
Qy	84	-----	83
Db	4944	VVEGPDVTHIAPGDRVMGVSGSPASFAVTDARRLTHLPEOCTWEIGASPLVFLTAYH	5003
Qy	84	-----	83
Db	5004	ALKELGLTAGKVLIHAGAGVGMAAIOIARHFGAEVFATASEAKMDVLSRLSGVADDDHI	5063
Qy	84	---SQT- :	87
Db	5064	ASSRTLDEFAAFVAVAGENGLDVVLNSLAGEFVDASHMLLTGGRFLEMCKTDIREDDSV	5123
Qy	88	-----	87
Db	5124	PDGITYQSFDLA FVDPEVIGMGRHETELFAAGDLHPLPVRAWDIRHATDAFRHMSMARH	5183

```
Qy 88 ----- 87
Db 5184 IGKIVLTPRAWSEGTVLVTGGTGGGLVARHLWMERGVRRLLLTSRSLDAAGAREL 5243
Qy 88 ----- 87
Db 5244 VAELENLGAESVAACDVADRDVDTLITAGIPAEHPLRAVHTAGVLDGVLGSLTEERL 5303
Qy 88 ----- 87
Db 5304 ATVLRPKVDGAWNLHEATRRLLDAFVVFSSVAGVFGGAGQANYAGNAFLDALTAHRQS 5363
Qy 88 ----- 87
Db 5364 LGLPGLSLAWGAWAOGAGTAGLBERDIRRAAESGMPLITVEQGLALLDALTGSAALV 5423
Qy 88 ----- 87
Db 5424 PVRDLAVLRARGTVAPLMRGLVRAPARPAATAATGDTALVDRLTRLORTERRDALLTL 5483
Qy 88 ----- 87
Db 5484 VREQAALVLHGSGGGIDPSRAFRDLGFDLSLTAVELNRNLGAATGVRLSATAVFDPYTV 5543
Qy 88 ----- 91
Db 5544 ALVDHLLTELLGPDAESDAEAPAEALLDDDPPIVIGMACRFPGGITSPEDLWRLLEG 5603
Qy 92 ----- 98
Db 5604 SDVISDFPNRGWDVDSLYDPAPHTGTSYTRSGGFLHNAADFADDFGMSPREAMATDS 5663
Qy 99 ----- 98
Db 5664 QORLLESTWEAERAGIDPVLGRSGTVGFAGVMYGYGTALDGEFEGQGSALS 5723
Qy 99 ----- 98
Db 5724 ASGRVSYTFEGEPAMTIDTACSSALVAMHLAAQALRSCECSLALAGGVTVMSIPTFVE 5783
Qy 99 ----- 98
Db 5784 FSRQGLSPDRCKPFSESADGVHSEGVGMULLLERQSDAIRNGHRVLAVVRGSAVNQDG 5843
Qy 99 ----- 98
Db 5844 ASNGLTAPNGPSQORVIRQALASGGLSTGVDVAEAGHTGTTGLDPIEAQALLATYQNR 5903
Qy 99 ----- 98
Db 5904 PEDRPLLGSVKSNIGHTQAAGAAGVIKVMAMQHGVLPRLNISEPSSHVDWSAGAVE 5963
Qy 99 ----- 98
Db 5964 LLTEQTEWETDRVRAGVSSFGISGTNAHVILEQPAKVIQGTVAASAAAPGDVVEPAV 6023
Qy 99 ----- 98
Db 6024 VPWVLSKTPPEARLDQAARLLATVTEBPAPRPVDVALSLATQRSQFAHRGVVLAGEHEET 6083
Qy 99 ----- 98
Db 6084 VRAALAALADGEPHSSVTGTVSSGRRALFSGGOSRLGMGRGLYGRFPVFAELDAVLA 6143
Qy 99 ----- 98
Db 6144 VLDGELGSLREVMWGEDAGLLNETGHTQPALFAVEVALYRLVESNGVRPFDVAGHSIGE 6203
Qy 99 ----- 98
Db 6204 IAAAHIAGVFLEDAARLVAARGRLMOALPAGGAMVAVQATEDEVPHLSGDVSIALLNG 6263
```

```
Qy 99 -----DLVLYI--TRFDLE----- 110
Db 6264 PTVSVVSGVEDAVLEVAARFEAEGRKTRURVSHAFSLPMLDPMLEDFRAVAEGLSPAAP 6323
Qy 111 ----- 110
Db 6324 VIPVSNLTGSLATAEELCSPEYVVRHVRBAVRFADGVAALAEAOGVTTFLELPDGVLSA 6383
Qy 111 ----- 110
Db 6384 MAQESLTGDETIVTIPLLRKORDEETAALTALGRLLHATGTRVDWAGFLAPAGARTVDLPTY 6443
Qy 111 ----- 110
Db 6444 AFQHRPEWPSGKRGATDVNSVGITAAAHPLLNNAVVELAEGEGLVFTGRLSUNSHWLADH 6503
Qy 111 ----- 110
Db 6504 AVMTGVLLPGTALLELALRAGDEVGCDHVBELTLAAPLVLTGOGAVRTOVRVGAEDAAGR 6563
Qy 111 ----- 114
Db 6564 RTVTIHSRPERSTDAPWTHQATGVLATGLPTAAAPPDATVMPADAEFVDLTGTFYEAREA 6623
Qy 111 -----LPDG----- 114
Db 6624 EGPAYGOAFOGLKAAMRRGDEVFAEITLPDGTOKAESFGLHPALLDAGLHAAMLAAPTG 6683
Qy 115 ----- 114
Db 6684 DGTGTVFPFSGMNGVSLVAGASSVVRVRLRDTGDTLTFLAVADTTGAPVAAVQALTMRAVS 6743
Qy 115 ----- 114
Db 6744 AEALSSAAALVRDSLFRDLMTPTVPAGSAPTAALVGEDPGLVRLREAGTEIGRRASL 6803
Qy 115 -----NRQV----- 118
Db 6804 TDDAELPDLLLPVTABEEHEAADTPAVTHALTQVLELVRARLEDERHAATREFVFTRG 6863
Qy 119 ----- 118
Db 6864 ATTGHDLAAAAMVGLVRSASENPGCFGLIDLDPETTRPLPLTALLGDEPOLRLODDLR 6923
Qy 119 -----RGVTQL----- 124
Db 6924 AARLVRPAPATDAVSAGGEGAVLVGTGGLGAVLARHLVAEHGVRELVLVSRGGAA 6983
Qy 125 ----- 124
Db 6984 AGAAELVAELAESGARATVACDVTDRAAVAELVAAPVSAVHVSAGVLDGLIASLTPE 7043
Qy 125 ----- 124
Db 7044 RLSGLVRPKVDAAMNLHEATRDLDLDAFVVFSSVAGVFGNAGQANYAAGNAFLDALMEHR 7103
Qy 125 ----- 124
Db 7104 RAAGLPGLSLANGPWEQTGCGMTGTLSDAELERLASSGVPPPLAVEOGLALFDAAALSTED 7163
Qy 125 ----- 124
Db 7164 AALVPVRLDLPALRGQGEVPPLLRSLIRASRRRAAVAGSATATGLRDLRLAVLDAVERREV 7223
Qy 125 ----- 124
Db 7224 LVDLVRGQVALVGLHADATAVHPTRAFRDLGFDLSLVELNRNLNAVTLGLRLPATWFDY 7283
Qy 125 ----- 124
Db 7284 PTVEVLVDFILDELGTVEAEVVGRTAAVDDPVIIVGMACRYPGGVASPEDLWRLVT 7343
Qy 125 ----- 124
```

Db	7344	EGTDAVSAPPVNRGMDVENLYHPDPDNICTSYTRSGFLHEAGEPDPFGMSPREALST	7403
Qy	125	-----	124
Db	7404	DSQORLLLETSWEATERAGIDPVGLRGSATGVFAGVMYSAMLASPEFEGFQSGSSP	7463
Qy	125	-----	124
Db	7464	SVASGRVSYTFGPEGPAVTDTACSSSLVAMHWAMQALRSGBISLALAGGVTVMSTPGVF	7523
Qy	125	-----	124
Db	7524	VDFARORGLSPDGRCKAFSDSADGVGMSEGVGMVLEROSDAIRNGHEITLAVVRSAVNO	7583
Qy	125	-----	124
Db	7584	DGASNGLTAPNGFSQORVIRQALASGGLSAQIEDVVEAHGTGTLGDPTEAQALLATYGR	7643
Qy	125	-----	124
Db	7644	DRDPEQPLMLGSVKSNIIGHTQAAAGVAGVIKMVMSMRHGVLPRTLHVDAPSHVDWTEGA	7703
Qy	125	-----	124
Db	7704	VELLTQTEWPETEHVRRAVSVSGISGTHNAHVILEQPAKVIOGTVIGGSTPESGVVPS	7763
Qy	125	-----	124
Db	7764	VVPWVLSGKTPEALRSQAALLASVEAELDRPLVDVGSSILVAARSLFEHRAVVLATDADT	7823
Qy	125	-----	124
Db	7824	AARALAALVGEPPAAVSGPARTGRSAALFSGQSGORLGMGRELYGRFPVFAEALDAVL	7883
Qy	125	-----	124
Db	7884	AVLDGELDGLSLREVMGEDAGLLNETGWTQPALPANEVALHRLVSEFVTPDFVAGHSIG	7943
Qy	125	-----GGA-----	127
Db	7944	EIAAAHTAGVFSLEDAARLVAAARGRLMQALPAGGAMVAQATEDEVIPYLSDEVSAALN	8003
Qy	128	-----	127
Db	8004	GPASVVYSGAEGAVLEVAARFEAEGRKATMLRVSHAFSPMLDPMLEDFRAVAEGLSFAA	8063
Qy	128	-----CSP-----	130
Db	8064	PVIPVNSLTGSLATABELCSPEYVVRHVRFAVDFADGVTTLEEQVTTLELCPDGVLS	8123
Qy	131	-----TWSCLE-----	135
Db	8124	AMQESLTGDEAVTVPLLRKNTAEAAATVAALARLHIGGLRIDWSALFAGTGARRVELPT	8183
Qy	136	-----	135
Db	8184	YAFQHWFWPAAPLGGDVRAAGLSAEHPLLGAAVELAAGEGVLTGRLALHSHPLWAD	8243
Qy	136	-----	135
Db	8244	HTVDTGTVLLPGTALLELAIRAGDEVGSDQVEELTITAPLVLPERGAVQVQVAEAPDAAG	8303
Qy	136	-----	135
Db	8304	RRVGVYARPEGTGTANSQHAAGVLAAPAGSRPAPAPADTAVMPPEGAQAVDADGCVEEF	8363
Qy	136	-----	135
Db	8364	AAAGFGYGPLFOGLRAAARRGDEWFAEVALPEGTTAAGTGLHPALFDSALHALLLTPA	8423
Qy	136	-----	135

Db	8424	ADGOTAGLPESWTDVSLHASGATVLRVRLTPTGDRTLSLSAVDPSGTPGVVSGSLVTRSV	8483
Qy	136	-----	135
Db	8484	DGLTTCTEALLRDADFFGVETWTPAQGSEAAATVALLGTGDPGLGAAALTGAGTTVRTY	8543
Qy	136	-----	135
Db	8544	PDVAALVAATVAADGPVPEVVVHVASHDGEPAASDAARHLTADALDTVQSRLSRLSLAGA	8603
Qy	136	-----	135
Db	8604	RLVEVTRGAVDTRDLGAAAVQGLVRSQOTENPGALGLVDLDTGPESAAALAAALGSAEPQ	8663
Qy	136	-----	135
Db	8664	LALRDGRPHARLARLARPADTAAPTAWQSDGTVLVTGGTGLGCAQFARHLVDAYGVNRL	8723
Qy	136	-----	135
Db	8724	LLVSRRGDAPGTTTELVAELIAHGAEVTVQACDVADPAVTALVAGIPAEHPLTAVVHTA	8783
Qy	136	-----	135
Db	8784	GVLDGVTIGSLTGERLARVLKADAAWNLHQATRGDLDDAFIVPSSVAGVFGGAGQANY	8843
Qy	136	-----ITE-----	138
Db	8844	AAGNAFLDALMEHRRRAAGLPGLSLANGPMDQAGGTMETLSDAEAEARLARAGMPPLAAROG	8903
Qy	139	-----	138
Db	8904	FALFDAALAAGNGTGTGSEDTTGTALVVPVRLDLAALAAOGEVPAVLRLVTRTRRTVA	8963
Qy	139	-----DTGFDL-----	144
Db	8964	GGTVTADGLVARLTALTAEERREALLGLVRTQAALVLGHADAASVDAQAQFQDVGDSLIT	9023
Qy	145	-----	144
Db	9024	AVELNRLSAAATGLRLTATLVFDYPNAATLASHLHDELFGAETESALALPVKQLPATTTDD	9083
Qy	145	-----	144
Db	9084	PIVIVMACRYPGGVASPEDLWRLLEGTDAITEPPVNRGWDIESVYVHPDPEHLGTSYTR	9143
Qy	145	-----	144
Db	9144	SGGFLHEAGEPDPFGFMSPREALATDSQORLLLESSMEAIERAGIDPVGLRGSATGVFA	9203
Qy	145	-----GVTI 148-----	144
Db	9204	GVWYSDYGSILGKEFEGLOGQSAGSVASGRVSYTFGPEGPAVTDTACSSSLVAMHILA	9263
Qy	145	-----	144
Db	9264	AQALRTGECSLALAGGTV 9282	

RESULT 8

Q93NX9 ID Q93NX9 PRELIMINARY; PRT; 9510 AA.
AC Q93NX9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE AmphI.
GN AMPHI.
OS Streptomyces nodosus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=40318;
RN [1]

RP SEQUENCE FROM N.A.
 RA Caffrey P., Lynch S.V., Flood E.M., Finnian S.M., O'Liynyk M.:
 RT "The amphotericin biosynthetic gene cluster from Streptomyces
 RL nodosus";
 DR Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF357202; AAK73501.1; .
 DR InterPro: IPR001227; Ac transferase.
 DR InterPro: IPR002198; ADH short.
 DR InterPro: IPR004410; FAD.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR000794; Ketoacyl-synt.
 DR InterPro: IPR001005; Myb_DNA_binding.
 DR InterPro: IPR003880; Ppantne_attach.
 DR Pfam: PF00698; Acyl_transf.; 6.
 DR Pfam: PF00106; adh_short; 1.
 DR Pfam: PF00109; ketoacyl-synt.; 6.
 DR Pfam: PF02801; ketoacyl-synt_C; 6.
 DR Pfam: PF00550; pp-binding; 6.
 DR TIGRFAMs: TIGR00128; fabD; 6.
 DR PROSITE: PS50075; ACP_DOMAIN; 6.
 DR PROSITE: PS06066; B_KETOACYL_SYNTHASE; UNKNOWN_6.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
 DR PROSITE: PS00012; PHOSPHOPANTHEINE; UNKNOWN_6.
 KW phosphopantetheine.
 SQ SEQUENCE 9510 AA; 992817 MW; 1A8E003A15D478CF CRC64;

Query Match 31.2%; Score 241; DB 2; Length 9510;
 Best Local Similarity 1.0%; Pred. No. 18;
 Matches 95; Conservative 18; Mismatches 31; Indels 9162; Gaps 26;

QY 3 GGILH----- 7
 DB 88 GGFLHDAPEFDFFGISPRAVAMDPOORVLESANEAFERAGINPTSVKSGRTGVIG 147
 QY 8 ----- 7
 DB 148 AMAQYRVCADAGEFQLTGNTGVSLSGRISYFTGTGPAVTVDTACSSSLVAVHLATQ 207
 QY 8 ----- 7
 DB 208 ALRAGECTLALAGGVTVMSGPGTFIEMRGGLSVDRGCRSFGDTADGTGAEGVILVL 267
 QY 8 ----- 7
 DB 268 ERLSDAIRGREILAVRGTAVNQDGSNGLTAPNGPSQOAVIEQALYNARLSAGDIDV 327
 QY 8 ----- 7
 DB 328 EAHGTGTLGDPVEAQAALLATYQQORDEKPLLLGSKVSNISHTQAAAGVAGVVKVMAM 387
 QY 8 ----- 7
 DB 388 RHGVPLRTLLADEPFRHVDMSGAVRVLTENTEMPATGAPRAAVSSFGISGTNAHTIVE 447
 QY 8 ----- 7
 DB 448 QAPEPEPADPEDDAPSTPAAVTGVLPVLLSGRSPEVLRAQAAALLTLTGTPPPADLA 507
 QY 8 ----- 7
 DB 508 YSLATRTAFEHRAVILLASDLPPELTGRLTAIAEGTDPVLAADVGTGTARTETRLAVLFTG 567
 QY 8 ----- 7
 DB 568 QGAORLGAGRELAARFPVFAAALDAALDAFTPHLDVLRKVLWGEDADRLDRTTEYAQPAL 627
 QY 8 ----- 7
 DB 628 FAVEVALYRLLESFEVVKPDHLACHSVGEIAAAHVAGVFSLLDDAATLVAARGRLMQALPEG 687
 QY 8 ----- 7

DB 688 GAMVAVQASEDEVAPLLACHEDLVSLAANGPSAVVLSGDETTVTVELAARLAADGRKTSR 747
 QY 8 ----- 7
 DB 748 LRVSHAFHSPMLMAPMLDEFNNVVEGTLHSPLLPVVSDVTGEPATVAQLTSPDYWDHVR 807
 QY 8 ----- 7
 DB 808 QAVRFADGIDMLARHDVTAFLGLGPDVLSAMAQNCIDAGSDALTVPALREGRPEDHTF 867
 QY 8 ----- 7
 DB 868 TAALAAHTGCTALHWDACFTGTGARRTDLPTVTFQRRRYWPRAVOGGAADLRSLVGLAA 927
 QY 8 ----- LELLY----- 12
 DB 928 HHPLLSAAVSLADSEGALLTGRISSLSHPMLADHTVRGATLLPGTAPLELAVRAGDEVGC 987
 QY 13 ----- 12
 DB 988 DRVDELTLAAPLVLPQGGVQVOLWIGNPDASGRRSVTVYGRPDADADAPWTSHTATCVLS 1047
 QY 13 ----- 12
 DB 1048 ASRTSDFDATVWPPADAETLPVDGLYERLAEGGFGYGLFQGLRAAARRGDEVFAEWL 1107
 QY 13 ----- 12
 DB 1108 PESHTDAESFGLHALLDSALHAASFPVLDERRAAGLPFSWEGVLSHAGATTLRVLA 1167
 QY 13 ----- 12
 DB 1168 PAAGDAVAIAVADDGQVLVSADSLILRAVAAREIDAAAALVRDALFRLDWVPVTAAS 1227
 QY 13 ----- 12
 DB 1228 GTAALVGEDPFGRLALPQFGLAVHPDLADLAAADGAVPDTVLLPLTGTGPDADPTAA 1287
 QY 13 ----- 12
 DB 1288 HRAATEALAAVRTWLEQDERFAASRLALVTRGATTGHDPAAPAAVWGLVRSQAENPORFL 1347
 QY 13 ----- 12
 DB 1348 LVDLDADQDTPALPAAALTSEEPQAVRGELRAARLVRPASTAEAVPAGGEGAVLVT 1407
 QY 13 ----- 12
 DB 1408 GGTGGLGAVLARHLVAEHGVRELVLVSRGGAAGAAELVAELAESGARATVWACDVTR 1467
 QY 13 ----- 12
 DB 1468 AAVAEVAAHPVSAVHVSAGVLDDGMVGTLLTPERLTTVLRPKVDAANLHEATRDLDKA 1527
 QY 13 ----- AVGP----- 16
 DB 1528 FVLFSVAGVLGSPQANYAAGNAPLDAALAAHRAAGLPGLSLANGPWEQTGGTGGISE 1587
 QY 17 ----- 16
 DB 1588 DDLRRMARAGTPALTVEOGLALLDGDALDGDALAPVRLDLSVLRQAGEVPPLLRLSLIRG 1647
 QY 17 ----- 16
 DB 1648 RSRRAAVAGSATAGLAQRLARLDAESRDELVLDIRGOVALVGHATGAIEIDAGRPRE 1707
 QY 17 ----- 16
 DB 1708 LGFDSLTAVELNRNLNTVTGLRLPATLVFDYPTVSHLSYVLDELGLTTEVEAEVVQRCTA 1767
 QY 17 ----- 16
 DB 1768 AVADDPVIVGMACRYPGGVTSPEDLWRLVTEGTDVAVSGFPVNRGMVDVENVLYHPDPDHPG 1827

Qy	17	-----	16	Qy	30	-----	29
Db	1828	TAYTRSGFLHEAGEFDPGFFGMSPREALATDSQORLLLEASWEAIERAGIDPVGLRGS	1887	Db	2908	DGEQVALRATGVHGRRLVRHIVDELPSADQFTASGVLITGTTGGLGAETARMLARSG	2967
Qy	17	-----	16	Qy	30	-----	29
Db	1888	TGVFAGVMYSYDYSAMILGSPFEFGSGSSPSLASGRVSYTLGLEGPVAVTVDTACSSSLV	1947	Db	2968	AHLVLTSSRRGPDAPGAELRAELRQSGASVSIVACVADRDLAAVLDDGLSADOPLTGV	3027
Qy	17	-----	16	Qy	30	-----	29
Db	1948	AMHWAMQALRSGEISLALAGGVTMSTPAVFDFARQGLSPDGRCKAFSDSADGVGWE	2007	Db	3028	HTAGVGHYGLDALTPAEFACLTAAKLAGAAHLNLLDRELDFFILFGSIAGVWGSDQ	3087
Qy	17	-----	16	Qy	30	-----	29
Db	2008	GVGMLVLEROSDAIRNGHOILA ^W RGSAVNODGASGLTAPNGPSQORVIRQALASGGLS	2067	Db	3088	SAYGAANAYLDALALARRARGLAATSIANGPWGTTGMAADDAVSGTLRROGLGLDPAPA	3147
Qy	17	-----	22	Qy	30	-----	29
Db	2068	AGVDVVEAHGTTGLDPIEAQALLATYGRDRDPEOPLLLGSKNIGHTQAAAGVAGV	2127	Db	3148	LTEMRRVVRQDVTTVADVDWTRYAPLFTSARPSALISOLPEVRALAAENTPADTGAS	3207
Qy	23	-----	22	Qy	30	-----	34
Db	2128	IKWMSMRHGVLPRTLHVDAPSSHVDWTEGAVELLTEQTAMPETGRPRAAVSSFGISGT	2187	Db	3208	EIVQVRSLSEPEQLRLTDLVRTEAATVLGHSSAGAPVEDRAFREIGFDSLTAVELRKH	3267
Qy	23	-----	22	Qy	35	-----	34
Db	2188	NVHTVLEQAGTTVPAPAPERTAGAVPLLLSGRTRDALRAQAARLLTHLQNHPEPSLAD	2247	Db	3268	LGAATGSLSPSTMVFDYPTPLELAQYLAEMVGSVLEVAGPVATGTTDDEPIAIIIGMSCR	3327
Qy	23	-----	22	Qy	35	-----	34
Db	2248	LGHSLATRRSRFERRAAVIAODREGLLASIGSLAAGRPDPANVEGEAAGRAVAVMFSQ	2307	Db	3328	YPGGVSPPEQLMDLVLSGTDAITDPPVNRGWNATAGLDPPDPDHPTTSTOGGFLHEADE	3387
Qy	23	-----	22	Qy	35	-----	34
Db	2308	GSQRAAMGRELYETQPRFAAFDEVCAALDPLDRPLREVVFAAEGSEEAALLDRTGWTQ	2367	Db	3388	FDPMFEGISPREALVMDPQORLLLETTWEAFERAGLTPTDLRGSLTGTFIGSSYOEGWG	3447
Qy	23	-----	22	Qy	35	-----	34
Db	2368	PALFAVEALYRLVESWGRADFVTGHSIGEIAAAHAGVFTLQDAARLVAARATLMEAL	2427	Db	3448	AGDGAEGHLVTGTSPLSGRLAYVFGLEGPATVDTACSSSLVALHLACOALRNGESNL	3507
Qy	23	-----	22	Qy	35	-----	34
Db	2428	PSGGAMVAVQATEEVEAPLLGEGLSVAANVGPTSVVVSGDEDPAVELAAEFSGRGRTRK	2487	Db	3508	AVAGGATVMTTNPNAFVAFSRORALAOGRCKAFSESADGMTLAEGVIGVILVERLSDARN	3567
Qy	23	-----	22	Qy	35	-----	34
Db	2488	LRVSHAFHSPHMDAMLDARTVAETLSYAAPRIPLVSDLTGRRADDAEVRTADYVVRHVR	2547	Db	3568	GHPLAVIRGSAINQDGASNGLSAPNGPSQORVIRQALANARVAPGEIDLEAHGTGTPL	3627
Qy	23	-----	29	Qy	35	-----	37
Db	2548	EAVRFADCVRTLRDAGATFLFELGPDGLLTAMAE ^{QEDT} TLGDERYDHTALVPLLRADRPES	2607	Db	3628	GDP ^{QEDT} IEAQUALPATYGRTRTPETALLGSKVKNIGHGSSQASAGVASIIKMMNALRHGVMPOTL	3687
Qy	30	-----	29	Qy	38	-----	37
Db	2608	AAATAAARLQIHGVLDLNTAYLAGTCARRVDLPTYAFOHAHYWQLPSPAAPSAGDPADQ	2667	Db	3688	HADEPSSHVDWSPGTVRLIGENTDWPOTGRPRAAVSSFGISGTNNAHVILEQETEAPAAE	3747
Qy	30	-----	29	Qy	38	-----	37
Db	2668	KLWAAVERGDAEAALVGLDEDSLTPLDSLPLALSSWRRGNQEKALLDTLRYRVEWTRL	2727	Db	3748	DEQLAPAPLPAAGVVPWLLSARGAALREQADRLLTLVTADPAARPIDIGLSLATSRA	3807
Qy	30	-----	29	Qy	38	-----	37
Db	2728	SKPAAVLDGTWLVSSDATADDETLDDGLAEALGAHGARVRRLVLADCADRAVLGAR	2787	Db	3808	LFEHRAVVPAGTDPLEALRAVAADGPSGVVARGVADVAGRTVFVFGOGSOWAGNGAQ	3867
Qy	30	-----	29	Qy	38	-----	37
Db	2788	LADTENADNTAOVLSVPLDERPTDGPAGFTOGLALTIALVQALADTGHGRLMTATRGA	2847	Db	3868	LLDESPVABERIAECAALAEFTDWNLIIDLVRGAEGAPTLE ^{QEDT} RYDDVVQVPASFAVMVSLAAV	3927
Qy	30	-----	29	Qy	38	-----	37
Db	2848	VSTGPADPVTHPAQATAMGMGRGVALEHPRLWGLGLVLPADFDRGAGORLAEVLAVKADP	2907	Db	3928	WRAQGVPEDAVVGHSQGEIAAAVVVSGALSRLDCGARVVTLRAQAIGRSLAGRGMMVSVALP	3987

Db	3988	VAEVEARLEAFEGRVSVNAENGPRSSVVAGEPEALDELHAQLTAEERARRVAVDYASHS	4047
Qy	38	-----	37
Db	4048	PHVEDLHDEILLELAEVAPRTSEIPFFSTVTGDMWDTTMDAGYWRSLGRVLFADAVR	4107
Qy	38	-----	37
Db	4108	DLIAADHRAFIENVSSHPVLAMSVQMDIDDAGVAGVASCGLRRDNGGLDRFLLSAAEVFVR	4167
Qy	38	-----	37
Db	4168	GVQVDAAVFEFGTASRVLDLPTAFQHENLWAMAAAPEAVTAADPEDAAFTAVEDGDVS	4227
Qy	38	-----	37
Db	4228	ALTAALGTDEDSVAAPALSSWRRARKERSTVDSWRYRPTWKPVTKLPQRTLDGTWLLV	4287
Qy	38	-----	37
Db	4288	SADGVDDTDVAEALETGAEVRRLLVDESCTDRAVLRLRLTDADGLTGIVSVLAGAERTG	4347
Qy	38	-----	37
Db	4348	AVPGTGLVLGVALTVQALGDAGIDTPLWALTRGAVSTGRSDKVTAPOVQVGTIGWT	4407
Qy	38	-----	37
Db	4408	AALECPGRWGVVDPETLDRAGORLAVALAGALGDDDDQIALRSSGVFTRRIVRADAAP	4467
Qy	38	-----	37
Db	4468	DCSARDWKPRGTTLVGGSGTLAPHLARWLAEQGAHLVLSRRGPEAPGAELRAELAE	4527
Qy	38	-----	37
Db	4528	RGTEITLAACDITDRDAVAALLESUKAEGRTVRTVVHTAATIELHTLDATLDDDFRVL	4587
Qy	38	-----	37
Db	4588	AKVTGAQILDELLDDELDFFLYSSTAGMMGSAHAAYVAGNAYLAALAEHRRARGUTA	4647
Qy	38	-----	37
Db	4648	LSLSWGIWADDLQGRVDPQMIIRRSGLFEMDPQLALSGLKRALDDDEQVIAVADVWETY	4707
Qy	38	-----	37
Db	4708	HPVYTSARPTPLFDEVPEVQRLTAAAEQSAGDPARGFAAALLALPAAEQHRKLLTETVRT	4767
Qy	38	-----	37
Db	4768	EAASVLGSSABDLTDQAFRDVGFDLSLTVAGLRNRLASVTGLTLPSTWTFYPNPAALA	4827
Qy	38	-----	37
Db	4828	GFLHSELADVHSAGAVATAGAPVDDDDPIAIVGMSCRYPGGITSAEQLWRVSLVEEDAVS	4887
Qy	38	-----	45
Db	4888	VFPADRGWDAEALYDPPDASGRYSVQGGFLRDVADFDGFFGISPREALSMDPQORLL	4947
Qy	46	-----	51
Db	4948	LETAMEVFENAGLDPVQGRSRTGTFIGASYQDYGAAPVSGSEGCHMITGSLPSVLSGR	5007
Qy	52	-----	54
Db	5008	VSYLFQLEGPVTLDTACSSSLVAIHLACQSLNGESTLALAGGASINSTPMSFIGFSRQ	5067
Qy	55	-----	54
Db	5068	RALAEDGRCKAVBAGADGMTLAEGVGLILLERLSDARRNGHEVLAVIRGSVNVQDASNG	5127
Qy	55	-----	54
Db	5128	LTPNGPSQORVTRQALANAGVEANDIDVLEHGHTGTALGDPIEAQALFATYCKDRDPER	5187
Qy	55	-----	62
Db	5188	PVLGSKVSNIGHTOMASGVASIIKLVLHALREGVAPKSLHIDQPSHVDWSSSTIOLLTE	5247
Qy	63	-----	62
Db	5248	RTEWPTGRPRRAAVSSFGLSGNTVHTVLEQAPAADAPAAEDTPAPRDALVPVLVSRGE	5307
Qy	63	-----	62
Db	5308	AALRAQAGALLDLLAERPPIHTDPLAFSLATSRAALEHRAAVVADDHEALVRLGTALROG	5367
Qy	63	-----	62
Db	5368	LPGAGLVQGRTRGRTAFLFTGQSQRGLGMRGELYERHPVFADALDAVLARIDGTTTERPL	5427
Qy	63	-----	62
Db	5428	RDVLFAAEGSQDAALLHRTGYAQPALFALEVALFRLLLESWGVTDPYLAGHSVGEIAAAHV	5487
Qy	63	-----	65
Db	5488	AGVLDLDDACTLVAARGRLWQALPEGAMVYLAEEDEVLPHEGLADQVVAANGVPRS	5547
Qy	66	-----	69
Db	5548	VVAGEEPPVLAAAHFAEQGRRTKRLRVSHAFHSPMLDPMDDFAAVARALTYHAPSIP	5607
Qy	70	-----	69
Db	5608	FVSNVTGLAAPOVCTADYVWVSHVRSVAFADGIGHLSTGGVQTFLELPGDGVLSGMA	5667
Qy	70	-----	69
Db	5668	RESLTDASRTALLPTLRGDRPEEQALVTAVAHAHAGFDVDTAWFOGSGARRVALPTYA	5727
Qy	70	-----	77
Db	5728	QFERYWPDTTAAGITAPAPGSALDAEFMAAVEHADVASLTSLSGLDDATVTAMVPALTA	5787
Qy	78	-----	77
Db	5788	WRQRGQSALDSWRYRVTWKPRGGAPGAAPTGRWLVLPFAEHRDEATAAADAADVEAALA	5847
Qy	78	-----	77
Db	5848	TATVREVTGTDRAALAARLTEAADGDTFQGVLSLLALAPGDAGHPCGAPALTLTATALQ	5907
Qy	78	-----	77
Db	5908	ALGDARIDAPLWNITRGAVAVGRSEQVTAPEQAAVWGLFRAAALELPAVGGSVDDLPEDL	5967
Qy	78	-----	77
Db	5968	DTQAARLRGILAAADGEDAVAVRASGVFLRLLAHSPAADTVGSAPDPAAGTVLITGTG	6027
Qy	78	-----	77
Db	6028	GIGGHLARLARDCAAHLLTTSRRGPDPAPGAGELRAELEESGARVTIAACDAARDALAA	6087
Qy	78	-----	77
Db	6088	LLATVPEDAPLTAVFHTAGVVDDHVDDELTPESPATVLAHKTVAARHLHDLTAGHDLAA	6147
Qy	78	-----	77
Db	6148	VLFSSTAGLGAAGQGNVAAANSQDLDALAEQRRAGQLPALSVANGPWAGSGMWADAABIA	6207

QY	78	-----	77	QY	86	-----	85
Db	6208	ARVRGGPEPLDPASCVAALLRAVDNGDTSAVADIDWERFLRAFASARPLPLVADLPET	6267	Db	7288	TASLGLOGDVTVTMVPAWSARRRTDARSTARDWRYHESMTALCTPAHSAGAGRVLALVP	7347
QY	78	-----	77	QY	86	-----	85
Db	6268	AGANAPAAAAGTSGUREQAGLPAERHAHVLDLRTQVAAVLGHSRDARTVDDDLAFRD	6327	Db	7348	AEHAGTDWAETLVAALGADPLVVDGTSGLARELADLVPOQATFAPRRRPARTLAAPAES	7407
QY	78	-----	77	QY	86	-----	85
Db	6328	LGFDLSLTVLELRNALNLTGLSLPASLVYDHTPREMADFLLAELGSLPESTRHAVATR	6387	Db	7408	PSTSTGTSGRHTESTHRTPLLDGOTSPAGTLPAGWTVVSVLLAAGTGTLPAADAPAA	7467
QY	78	-----	77	QY	86	-----	85
Db	6388	AVDEDPVAVGLACRFGGVSTPEELWQLLAEGRDGITGPPDRDNWDIAALGAGASDTLH	6447	Db	7468	VLEALEAAGVDAPLWCVTRGAVSVAGEAPAAVGOAALMGWRVALEHPERFGGLADLAP	7527
QY	78	-----	77	QY	86	-----	85
Db	6448	GGFLAQVADFARFFGISPREALAMPQORLLLETTWEALERAGIDPAALRGSTTGTVFG	6507	Db	7528	DADAATAALLAHLAEPGGEDQIAVRATGLFRRLVRTAVAPGDSGMRPHGTVLVVGTR	7587
QY	78	-----	77	QY	86	-----	92
Db	6508	TNGQDYNLLRRSTDVSGYVATGNTASVMSGRLSYALGLEGPATVIDTACSSSLVALHW	6567	Db	7588	ANGARAARWLAREGAARLVLTTSPADSATDTEELRAELGRGAETVAPYDGGORDAARA	7647
QY	78	-----	77	QY	93	-----	92
Db	6568	AGHALSAGECDLVVAGGVSVMSTPDSFVFPSTQGLAPDGHCKPFSDDADGTAWSEGVGI	6627	Db	7648	LLDGLTGLTAVVYADDTPADGPAAALAPVDTLAEATGRSLDAFVLFGSVAGVMGVRGT	7707
QY	78	-----	77	QY	93	-----	95
Db	6628	LVLERLSDAIRNGHEVRGIIRGTAVNQDGASNGLTAPNGPSQQRVIRQALADARLAPSEI	6687	Db	7708	DEAAEGYVDALARALRAEGTPALAVSNWMAELTDPSTTRHLRMNGLPVMDADAALTAL	7767
QY	78	-----	77	QY	96	-----	95
Db	6688	DAVEAHGTGTLGDPIEAQALLNTYGPGRDTPQLLGLTVKSNIGHTQAAAGVAGVIKML	6747	Db	7768	AGAVADGSAANTVADVVRWDTFAPAHETRPTALPDGLPEARTALAGADRAREGATASDG	7827
QY	78	-----	77	QY	96	-----	109
Db	6748	LAMQHGTLPRTHVTSPTSHDWSSGAVSLUTEERDWPETGRPRRAGVSAGVSGTNAHV	6807	Db	7828	YGRWILLEQPAARDRIILLALVSEKAALVLGHADTTLVEPDLPFRDLGFDLSLTAVDLRNLQ	7887
QY	78	-----	77	QY	110	-----	111
Db	6808	IVEQAPIDEPVVEPTEEPAPAAEVTTVPWIVSGRSREALQDQVDRLTAYAAARPELSPLD	6867	Db	7888	TAATGTLTPATLVFDHPNPAALAAHLRTELLGGATDTTAAPVAAPAAATGDDPIVIGMACR	7947
QY	78	-----	77	QY	112	-----	111
Db	6868	VGRSLATDRTLFPHRAVLLAGPDGVREARSVASRTRGRTAFLFSGOGAQRALMGRELYE	6927	Db	7948	FPGGVNSPEDLWQLVLDEVDVAGDFPADRGWDLDALAGDPCRSATDOGGFLYDATDFDP	8007
QY	78	-----	85	QY	112	-----	111
Db	6928	RYPAFADALDTVLAQFDTALDLSLEVLFABPGTPEABRLNETGTQPALFAVEVALHRL	6987	Db	8008	GLFGVSPREAMVMDPOORILLEASWEALERAGIDPGLRSGTGTGVFGGSGDYRPPAE	8067
QY	86	-----	85	QY	112	-----	111
Db	6988	VESWGLTPDQVAGHSIGEIAAAHVAGVFSLEDACALVAARASLMQELPRGAMLAVRATE	7047	Db	8068	YQWQTAQSASLLSCLRLAYTFGIQGTFSVDTACSSSLVALHLAAQALRSGECSTALAGG	8127
QY	86	-----	85	QY	112	-----	115
Db	7048	EEVAPHLTERVSYAAVNGPSAVVAGDEAEVAIAAHFEAQDRKTRRLRVSHAFHSPLMD	7107	Db	8128	VTVMATPVGFVEFSAGGALS PDGRCRPFSEDANGTGWSEGVMLVVERLS DARRNGHRVL	8187
QY	86	-----	85	QY	116	-----	115
Db	7108	PMLTAFARVAESLTYHEPVIPLVSTVTGPAGAELLTFGYVVRHVRHRTVRFADGVRAALRE	7167	Db	8188	AVLRGSAINQDGASNGLTAPSGPAQORVIRQALANARLEAADVDAVEAHGTGTLKGDPIE	8247
QY	86	-----	85	QY	116	-----	115
Db	7168	DCVTFVEIGPDGQLTAAQQTLDSGDEAPATVVVALQRDRSEETTLLEGLATLHTQGA	7227	Db	8248	AQALLATYQDRERPVLVGLSKANIGHTQAAAGVGVIVKWLAMQHGLLPRSLYTENPSS	8307
QY	86	-----	85	QY	116	-----	115
Db	7228	GPDMTAMFAGTGHHRVELPTVAFQRYWPEPAAPGATGVQDPVDAAFWAAFAVEREDLES	7287	Db	8308	HVDWTAGNARLLTTLTPWPESEVRRAAVSSFCASGCTNAHLILEQPEPVROPVEGARPAE	8367
				QY	116	-----	115

```
Db 8368 DGSPVPVLLSGRSONALRAQARKLLAHVAGHPEETVTDLARTLATTRAFFEHRAAATAAD 8427
Qy 116 ----- 115
Db 8428 RDALIAGLTAVAEGTTAAHIVEHQVERTGKHAVLPSGGGSOOLGNRDLYERFFAFAPAF 8487
Qy 116 ----- 115
Db 8488 DAVLARLDGHLGASLRDVVWHGDOETLNETGHTQALFAFEVALYRLVESNGVTPDFVAG 8547
Qy 116 ----- 115
Db 8548 HSGVEIAAAHVAGVLSLDDACRLVAARARLMDLPRGGAMVAVEATEOEVAALTEAVSI 8607
Qy 116 ----- 115
Db 8608 AAVNGPTSVVLAGEADAVLAVALAERGRTRSLKVSHAFSHVLMDBMLAEFRKVAESL 8667
Qy 116 ----- 121
Db 8668 RYEAPRIAVSVNTCDVAGADELDCADYVVRHVRCVTRFADGVLTALAGRGVRLLELGP 8727
Qy 122 ----- 121
Db 8728 AVLALAEOSAPDVTISIQAQRNRDEVETLVAALGRHLVHAAGPRWDAFFDAAPGSMWLDL 8787
Qy 122 ----- 121
Db 8788 PTYAFQGRFWPDTLPTAPOGATAGEADGTDSAFMDAVAQEDFGSLESVLDVEGDALSK 8847
Qy 122 ----- 124
Db 8848 VLPALLDWRRTSETQDLSWRHRIWVKLTGAAMAHKPLTGTWLAVVPEGLGEDPWT 8907
Qy 125 ----- 124
Db 8908 TALDAGTRVRAEVGADGREAMAATLREAAEGTRFGGVVSVLLALRETGHGVPEGVAL 8967
Qy 125 ----- 124
Db 8968 TGTLLQALGDAGIEAPLWCVTRSAVANSRDRPRRPLQAAVWGLGRVAALYPRNWGLV 9027
Qy 125 ----- 137
Db 9028 DLPERADERAAAGLAVALAGLDGEDQVAVRGSAVLARLVPAPDRGGSGAWDPTGTVLIT 9087
Qy 138 EDTGF ----- 142
Db 9088 GGTGAIGAHVARRLAKDGVQHVLLSRRGAEPGATALRDELQTLGAGVTLAACDASDRG 9147
Qy 143 ----- 142
Db 9148 QLAADVLAIPDDRPLTAVLHAAGVLDGVI DRLTPERYQKVFRAKVTSAALLDELTRDL 9207
Qy 143 ----- 142
Db 9208 LAAPVLFSSASSAVGNPCQANYAANAVALDAEARRVQNLPAITSVSWGANGGGGMADGH 9267
Qy 143 ----- 142
Db 9268 GADDAARAGVAAMPDPLAVESLLRLVTGKEPTAVVAEVDLDRFASAFGGARPSALLREF 9327
Qy 143 ----- 142
Db 9328 PGYRETVAAPASPEGTLADRLAGMAPARRLDTVVDLVRTRAAQVLYGYPIDIVAAGERS 9387
Qy 143 --DLGV 146
Db 9388 FRDLGV 9393
Qy 143 ----- 4
RESULT 9
```

Q9NFS3
ID Q9NFS3 PRELIMINARY; PRT; 16215 AA.
AC Q9NFS3;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE D-titin.
GN SLS OR D-TITIN OR CGI915.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Y.Q.; Broadie K.S.;
RT "Characterization of Drosophila D-Titin gene."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AJ271740; CAB93524.1; --
DR HSSP; P56276; 1TLK.
DR FlyBase; FBgn0003432; sls.
DR InterPro; IPR002106; AATRNA_ligaseII.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR003600; IG_like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; ig; 50.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGC2; 15.
DR SMART; SM00410; IG_like; 34.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00179; AA_TRNA_LIGASE_II_1; UNKNOWN_1.
DR PROSITE; PS0002; SH3; 1.
KW Immunoglobulin domain; SH3 domain.
SQ SEQUENCE 16215 AA; 1841509 MW; 242C8765E00F7603 CRC64;

Query Match 31.2%; Score 241; DB 5; Length 16215;
Best Local Similarity 0.6%; Pred. No. 90;
Matches 90; Conservative 29; Mismatches 29; Indels 15640; Gaps 27;

Qy 1 AAGG----- 4
Db 99 AAGNALFEGRLRGKPKPFVTRKCAPLLESOKFRMSYNEATGDVSLINQIGPGEDE 158
Qy 5 ----- 4
Db 159 YTCARNQGEAICSVVIOPEGAMPALQIONLEKNIYNGYSYTSIEEPRVDTPEYR 218
Qy 5 ----- 4
Db 219 LLREVSFREAIRTRSGVEODSLSQELDRNGPAQAPQISQKPRSSKLEGSDAVFTARV 278
Qy 5 ----- 4
Db 279 GSNPKPRLTWFHNGQRLVASQKYEISYSSGVATLRVKNTARDGGHYTLAENLQGCVVS 338
Qy 5 ----- 4
Db 339 SAVLAVEPAETAETAYEPKVDVMAEQLEAGKALPPAFVKAFGDEITEGRMTRFDCRVGTN 398
Qy 5 ----- 4
Db 399 PYPEVFWLINGROVDDASHKILVNESGSHLMTITNTRLDAGAVOCLARKAGEAIEA 458
Qy 5 ----- 4
Db 459 QLVNLEKEQVAVQFVQRFSTMTVREGPITMSANAIGTPQPRITWQKQGVQISSTAERF 518
Qy 5 ----- 4


```

Db 519 VGIDGATCLEIPRTVANDAGHYQCTAQNIAGSTANRARLYVEVPEQPNYEORELNLP 578
Qy 5 ----- 4
Db 579 PTKVIEPEIPGPEIILYLRHVERAKPHLRPGBEDRVYPPQFIILQNVQOTEGGRVHME 638
Qy 5 ----- 4
Db 639 ARIEPVGDPTMVVWYLNPRPLAASARATSVFKFGFIALDLLSIMGHDGSEYMCVNTNAS 698
Qy 5 ----- 4
Db 699 GVAESRAILSVVORPSIEQSSQNPNSLOYNQINQLEDYSRYQRTESIDEQLNQAPFIRPLR 758
Qy 5 ----- TLHLE- 9
Db 759 DLGEFEEGKNVHFEAQVTPVNDPSMRVEMYKGLPITASSRITAFINFGVYSLNHLRA 818
Qy 10 ----- 9
Db 819 EDAGTYTVRAVNRIGEAIQSISIRVHSQVTDLGIPEQORYIEKVBELEDYRKSQRR 878
Qy 10 ----- 9
Db 879 HVQEAEEATAPQFKTPIQNQLDLREHAHAHFEARLEPVGDSMTMRVWLKDGQPLEASSR 938
Qy 10 ----- 9
Db 939 IITYHNFGYVALTIKOLTIYDAGTYTCRAYNAMQDQTTVAQLTVISKNEIVSESQHPGGL 998
Qy 10 ----- 9
Db 999 QKIQHLEDSSRYGRREEETIYITQAPRFLGPKGTTKILEGQRAHFEARVEPOSDLGLVI 1058
Qy 10 ----- LLVA- 13
Db 1059 EWHNGESITAAARIQTYIDFGYVALDISQVRAEDAGVILVVARNKLGEAQQAQTMIVET 1118
Qy 14 ----- 13
Db 1119 RSSIDTSSMHRGLYEKTONLENKPFVPOYDIBEISKSPFVFTPLSDPKPIHDGKNHIL 1178
Qy 14 ----- 13
Db 1179 ECRLEPMGDPTMRVWFHNGRPVTVGSRFRITYYDFGFVALDIIKATAADSGETVTRATNH 1238
Qy 14 ----- 13
Db 1239 LGTAHTSACVRVIDHTDVVTETQNEQSLQEQIQLLEDSSRRHHQEEDITIMQAPQFTRGLH 1298
Qy 14 ----- 13
Db 1299 NIETIEGTVNHLERCLQPVGDPSPMRIEWFVNGKPVKTGHRFRPAYEFDYVALDLLGCYAI 1358
Qy 14 ----- 13
Db 1359 DSGYVTCOARNQOLGEAVTSCSVRIIAKNLDILETONESGLQIOYLEDSTRHRSEFVDE 1418
Qy 14 ----- 13
Db 1419 VVNIRPRFLTHPKSLTNTREGGHAHFECKIEPVTDPNLKVWFKNRGPITVGHFRFRPIHD 1478
Qy 14 ----- 13
Db 1479 FGYVALDIVHLIAEDSGYVTCRAYNLIGSDETOVELOCRSEGIQIVTVTONEAGLEQIHYL 1538
Qy 14 ----- 13
Db 1539 EDRSRYTRREIDESTKQAPVTTLSLKNVEIKENQRAHFECLLIPVSDPSMRVWYHNUL 1598
Qy 14 ----- 13
Db 1599 PLKSGSRFTETNFGFVALDIMSTLPEDAGTYTCRAYNAVGEAITSAVAVVHTKKSIYLE 1658

```

```

Qy 14 ----- 13
Db 1659 SOHETALPRLOHLEDGSKRQORISVQDEFVSQAPVFTMPVRDVRVAENQAVHFARLIPVG 1718
Qy 14 ----- 13
Db 1719 DPKLTVELWRNGQPIEASNRRTTTMHDFGYVALNMKYVNPEDSGTYTCRAVNELQOAVTSA 1778
Qy 14 ----- VGP- 16
Db 1779 SLIVOSKTSIOLETOHEAAMHKIHQLEDHSRYQRRESEYTVTTAPVFTVKLIGPSNLVE 1838
Qy 17 ----- 16
Db 1839 QOSAHYECRIEYPDPNLKVWFHNGKPLSTGHRFRITYYDFGFAALDILTYYAEDSGEYT 1898
Qy 17 ----- 16
Db 1899 CRVTNNILGEAINSIVLNVTSRSSIIHETOHEALTKIOHLEDTSRPFORKTDEOFHAEHP 1958
Qy 17 ----- D 17
Db 1959 QFGRPLRNAKYNEGAPVHLEATLIPVNDPTMKVWYCNCRPIOTGHRKFTTYDFGFVALD 2018
Qy 18 VFOAHOEDTERYL- 31
Db 2019 ILYAHAEADTGYMCKAKNAIGEAVTTCVAVNTANKTLDLTDLAORLEKIROLETYAPPP 2078
Qy 32 ----- 31
Db 2079 KPVVEEKQKPIFLUTPLSNLEHLEKEGEHAHLECRVEPINDPNLKIENFCNKQLPTGHRY 2138
Qy 32 ----- 31
Db 2139 RTHDFGVVALDILVYGEDTGYICKATNOLGEAVNTCNVLRNRRSMILDTQHPDALE 2198
Qy 32 ----- 31
Db 2199 KIQKLESKVPNARTEVGDAPIPPHFTAELRGSTEIYEGOTAHFEAOVAPVHDPNLRIF 2258
Qy 32 ----- TNLNI- 36
Db 2259 YHNGKPLSASRPHITFDGVVSLDITHAVAEADAGEYSRAVNALQOAVSNTLNRVIRG 2318
Qy 37 ----- GAELLR- 42
Db 2319 TIISDTQHPGLEKIRKLESTAPHORQEPETPGTRQRPVFTOPLQONIDRINEHOTAHFEA 2378
Qy 43 ----- 42
Db 2379 RLIPVGDPNLKVWYRNEKIIEDSSRITKQHDGFGFVSLDISHIRKEDEGVYMCRAVNLPG 2438
Qy 43 ----- 42
Db 2439 EAVTTASMRVWSEASIQMDTOHPDSISRIHQLEKPLAPRPTPEPERLFEKPIFTOLLTGPS 2498
Qy 43 ----- DPSL- 46
Db 2499 ELWEGTHAHFEARVVPVGDPSLKFEPWINGVELQMSRLRTHDFGVFLDITAVVPEDA 2558
Qy 47 ----- 46
Db 2559 GYVMCRAYNAAGEAVSSTAMKVTKNSIDGOPLIPESWEAIRLKEAAMNRVPMFVDSPT 2618
Qy 47 ----- 46
Db 2619 QOAPVFTTHQSVDKLEHGOHVLLLEAQVEPRADPNLRIEWFKNGISLTTGSRIRSTFDFG 2678
Qy 47 ----- 46
Db 2679 LVTLSINGLRADDSAIYTCATNOVGEAVSTSSLKIEDRHLQAESLHPDSLPRIGELEA 2738

```

QY 47 ----- 46
Db 2739 PKGRPEAPEPTIETPFIHLNNIECKESDNVRECNVEPARDPTMSIENFYNGQPLQA 2798
QY 47 ----- 46
Db 2799 AAKFKSIYDFGYCALDLTNSAENSGVYTKATNSKSGATTGTLKCTGGKTMFLDTQHP 2858
QY 47 ----- 46
Db 2859 OGEAGLEAVQETEELANRYTSKTKPETOYPPPPVMTPLQAEFHLSEAQPIHLEANVEP 2918
QY 47 ----- 46
Db 2919 KEDPNLFIEWYFNGKLNHSGRFKMTSEFGFVTMDMIEVYARDQGIYTCAYNKAGEAFT 2978
QY 47 ----- 46
Db 2979 STTIFCSSKENIESTQHPKGAEGLEQIQDLEDSLKDGSKPEQDLGIPRFTTFEVNI 3038
QY 47 ----- 46
Db 3039 ADIGEGELAHFEANLIPVGDQSMVFEWFYNGKVLASHRVRTIYAFGTVALEVLTKIED 3098
QY 47 ----- 46
Db 3099 TGTYTCRATNKGHTAEISCNLECVDPKPCQKPRFTSHIQPLEGLKDGQSAHPECTLI PVN 3158
QY 47 ----- 46
Db 3159 DPLKLVEMYHNGKLMRHSNRINKTVSDFGVVLDISYLDHDSGEYVCRAMNKYGEDFTRT 3218
QY 47 ----- 46
Db 3219 TLNCGRGVGYDLSQPSLORIRELECPGQQAADTAPLVAEPKFITQIVDVTKLVEG 3278
QY 47 ----- 46
Db 3279 QSAHPEARLTPTDPLVVEWYFNGKLPHGHRFRTFDFGIVILDILCYEENSGVYEA 3338
QY 47 ----- 46
Db 3339 RARNKYGEDVTRASLCKASKSLILDSQLPRMGEGGLEKIANLEYSMVRTREETTEETKG 3398
QY 47 ----- 51
Db 3399 KAPVFTVPLENIENREGENAHFEARITPADDPKLVEMYWNGRPLKAGSRFRFTCDFGF 3458
QY 52 ----- 51
Db 3459 VILEISPVYPEDSGEYSCRAINEYGEAVTTATMKIQKRSIIMESQLPKMGEGTIDRIAE 3518
QY 52 ----- 51
Db 3519 LEGLSRSTEFVDDDTGKPPFITSPFDMVIGENALAHFECKRLOPINDPSMRVDWFHNG 3578
QY 52 ----- 51
Db 3579 KALWAGSRIKTINDFGVILEIAGCYQDSGLYTCKATNKGHEATVSKLVQKRGQIVM 3638
QY 52 ----- 51
Db 3639 EPOLPSNFRGTGTSLOKLEETHMKREELVTEDEQNPMPKFTTEIKDNLDPPEGGIHFDG 3698
QY 52 ----- 51
Db 3699 RVEPVGPTMRIEWFYNGHVMATGSRVHQLNDGFGTALDQVYIYARDSGEYTCRATNKGW 3758
QY 52 ----- 51
Db 3759 TATTSKVTCKGHNIVYESQLPEGMTSEKLELGRIPAEAPKVVEEVFGPPKFTTQIT 3818
QY 52 ----- 51

Db 3819 SVTVDEAEAVRFECQVBKTDPSLRVVEWYRNGKPLPSGHRYNIFDMGFVSLDILYVYGE 3878
QY 52 ----- V 52
Db 3879 DSGEYVCRAINNYGEDTRATVSCCKLPTILLQNOVPRGMRKSDALTQMEATIKKYTSEV 3938
QY 53 HL ----- 54
Db 3939 HLTEDDLFDPRKQPPRFVQTQIKEQLTLEMAVTKPECOLAPVGDPNMKVWFFNGKPL 3998
QY 55 ----- 54
Db 3999 HKNRFPQIYDFGYVAMNFGWYVPEDSGEYVCRATNLNKGQETRAIIKVSGKPGIYVSOL 4058
QY 55 ----- 54
Db 4059 PAHQSIDRIREMEASQWVPDEVPDAKPRTKPVFVSKLEPQTVESGDPARFCVRVTGH 4118
QY 55 ----- 54
Db 4119 PRPRVMNLINGHTVVHGSRYKLTNDGMFHLDPKTRQYDTGKVEIARNSVGESIAATTEL 4178
QY 55 ----- 54
Db 4179 KVARSDDYRNVLKNSPRPWYDELAAYQKEROENELEKVFDERKQVLSEOSSHTLKGVE 4238
QY 55 ----- 54
Db 4239 HLKPKQYKPTPDQWQNVKAKKSEYNNKLTLEBOLLKETNLRRDTHQYAIPEGKVVS 4298
QY 55 ----- 54
Db 4299 SSQAKGMAQSYEENLOEKTSTTEVQAAPPKGIAQPSSESVHGREVHMNQOQVKEIQGD 4358
QY 55 ----- 54
Db 4359 LEITRKITATETTEVEHKGTTIQRVWQVPKAPVTKKIOPCRVFENEQAKFEVEFE 4418
QY 55 ----- 54
Db 4419 GEPNPTVKWYRESPIQNSPDLOIHTFTSGKSLIIRQVVEDSAVFCVAENRGGTAKCS 4478
QY 55 ----- 54
Db 4479 ANLVVEERRRAGKGIQPPSFVTTIQSTTVATGOLARFADAKVTGTRPLDVYMLKNGMKIQ 4538
QY 55 ----- 54
Db 4539 PSIKFKMLEEDSVHTLLIIEPFAEDSGRYECVAVNAAGEARCDGDCIVQSBKPEKPTTP 4598
QY 55 ----- 54
Db 4599 GSEKAPHIVEOLKSQTVEEGSKVIFRCRVGDKPTPTARMRGENFVRKPSRYFQMSRQGEY 4658
QY 55 ----- 54
Db 4659 YQLVISEAFPDEGTYKCAENKLSIQTSQALKVRPIENLDAPPITALKDVSVTEGMP 4718
QY 55 ----- 54
Db 4719 AQFKTTVTGKVKATSVQWFREGQLIPETPDFQIMFDGNSAVLLIGTTVEEDSGIFTVRVT 4778
QY 55 ----- 54
Db 4779 SSTGOVSESSAKLTVKKRRISAFQRLRTIDSAEDSESSSSGRESAPESPHAFOPGQOQOF 4838
QY 55 ----- 54
Db 4839 GQFLGVNCGOHOGRSROKKPKVRSKSLQPATKVIPIHKKSSRPTGRSLDKGVFLPGFKP 4898
QY 55 ----- 54

Db 4899 EPVKSWEETINLKATPIEKKKPAPKLEAAKVVLKSIKTERDQOIMSLGATLEQIAGKT 4958
Qy 55 ----- 54
Db 4959 EKEAIPWITMREKLKAVESVQQQLNKFDLDEVYLQPLEGGIETEGOLPQQAQVEQVQRTK 5018
Qy 55 ----- 54
Db 5019 EIQRKSMESVEIMEMTDQDKLITQOONAKDLIPWKEMRQQLKSVQRTVKQIDKFKIEE 5078
Qy 55 ----- 54
Db 5079 VELRHLOAQQAITEEYGTGTAETVVMIDESSKGSISKVLRDRDEQLQYEDQSNIYKQKFI 5138
Qy 55 ----- 54
Db 5139 TTEDVINHVSEKLEAORLIREQAVNMRQOQORPQLOPLTSVEDTVISQTSERQKLV 5198
Qy 55 ----- 54
Db 5199 QOQSFIIEAQRQFVQVEDSQMMSLEEYEHQKIINORTQOEAFSWRQPREPQKFIQVEDS 5258
Qy 55 ----- 54
Db 5259 TLLHLOERHDTQEQLLQQQPVMMDRGRKKPDQPPQYVQPOEQRVKKEEFVEKPKTYBEMHD 5318
Qy 55 ----- 54
Db 5319 ELVEPTIEQOPVPVMMERGGKKPQOEKTFEADHDELVEPTVQOPEPVPVMMERGGK 5378
Qy 55 ----- 54
Db 5379 KVAQOQTVLSQEVVQTSQVVEQQIVBETKKTAVRRVIPPREDQKVEQVTLKPTPRPRPK 5438
Qy 55 ----- 54
Db 5439 EAVKAEIQLPLRSTRPVPQVEAQKAYEATDELTEPPIPOQPVMMERGGKKPKQKP 5498
Qy 55 ----- 54
Db 5499 QEEVTEIPKTLIAVDTLEEBVPKTEPOPQVLMARGOKKQKQDEQKELPKSLIAV 5558
Qy 55 ----- 54
Db 5559 DTIEEDLIPVQEPQVPLMERKKKQPOQDVIEEKLDAVPTKYKANDVLPDEPKVEE 5618
Qy 55 ----- 54
Db 5619 KPEPVLWQGGKKIPKSEPTVEVHPDEVDAQIETVVKDEMIVEEKRRIKKTKRPKSTKE 5678
Qy 55 ----- 54
Db 5679 VTEELFEQPEEISPEEVPQKEVIEEIBEIVEEKRRLKTKKPKLTOQVTEETPHEE 5738
Qy 55 ----- 54
Db 5739 IIESEEVQOEIBEIVEEKKKVKKPKTVAEKQLKEEBEIPTEETVEEBEETAEDQQLVV 5798
Qy 55 ----- 54
Db 5799 EESKKVKVKKPTGCTVEKTDVEELPGSEVPVEVPVEVPDEVAPBEELIEEQEEIVDQD 5858
Qy 55 ----- 54
Db 5859 EIOEQKRVKAKKPKTKTIEKTEIEEBEDQPEBEVLQEEIIGEOBEITERQKRVKSIKKP 5918
Qy 55 ----- 54
Db 5919 KVVVTEKTVDQTEQPEKPEESQAEVKEVTYTESPKPKPAPAEAKVQVEKISLKAPRK 5978
Qy 55 ----- 54
Db 5979 QRLLPKEQVEEVLKPKVKKIVAVSEAEQPEPTEPEVKEFAITTTEDILDVTKKRVKK 6038

Qy 55 ----- 54
Db 6039 KKPKTVAAEESTEPABETEEFEEBEOEQVEVPVEBIPERPOVKEVADERKTAPKXP 6098
Qy 55 ----- 54
Db 6099 RKEEIIKVEEVALKRVTRPKKELPOEATIEBVRUKPQRTSIIKPEEVKLBEVDLOHVEK 6158
Qy 55 ----- 54
Db 6159 KEDEIVQEBEKRTKRVKKPKKHEDLPEIPDAEPTQLEEAHEIELEKQKPEEDPOQVPMKR 6218
Qy 55 ----- 54
Db 6219 GEKKQPEEVLBEKKWPSSGKRRLPEQOPEEVQLKPIPSKPIEEQOKPEKAIPGQPLVPE 6278
Qy 55 ----- 54
Db 6279 EKPESEEELELEPLKLPEDKKPKPKAKKPKKKKPKLKKATPSVDEVSVEVAEPDEPI 6338
Qy 55 ----- 62
Db 6339 AEEDEVEMPVDDVKVAVSEDLVPEEVPTEETPEAKQKAHKRTKRLKEASVEGQPO 6398
Qy 63 ----- 62
Db 6399 LLEAAIAEIEKVEISQISQKTIITLLKKTEDTRPQFITTEQLIIELDVEDVRRDLEMVKT 6458
Qy 63 ----- 62
Db 6459 SNIIKKERVRVLDSDQLPELELITQKRIQEGIDKVADEELIEDOQLIONQOETTSEV 6518
Qy 63 ----- 62
Db 6519 IQGERKLVKKKKEIKPPRITEKLPRQCVPEEPTVLECKVEGVFPPEIKWYFNILLFA 6578
Qy 63 ----- 62
Db 6579 SEKYEITVMEQVAKUKIAKFTPSDVGVYTCBAKNBAGVATSRNTIILEKEQCVPOFTKP 6638
Qy 63 ----- 68
Db 6639 LKIEFIEEKQPERLKVTVTCQVTGKPNPEVKWYRGIEEVIPESETVQMPFYDEKTDGVALE 6698
Qy 69 ----- 68
Db 6699 INPTNEAVVYSVQAQNOGRAIGNANILSRVDEVPREILKAPTVTPLSAVVVPTGTGLF 6758
Qy 69 ----- 68
Db 6759 FEAKYDGLPRPEVKWMRNGREIIEEETIETTTTIIKVVNMTRKRTGKYEVWAKNKV 6818
Qy 69 ----- 68
Db 6819 GEAKSSGSVVVSDQKDEQIKPPRFIQPLEPKYFGEHEVAIIIEAIVSEPLSSFQMFVHN 6878
Qy 69 ----- 68
Db 6879 EPIKSSNEVRIVSQANKSTLLIENFQSKFVGPFCTCAENVGSGVSTATVNLIPOEEABE 6938
Qy 69 ----- 72
Db 6939 FESPRFVEELVQPEVMDGEALLLTCQVTGKPTPKVWYHNAKIKTENKETTISODLOGV 6998
Qy 73 ----- 72
Db 6999 COLQITEVPFNEGOVCVATNKIGKSVKTNVKIQAFEIPDSSEITGIGTSEEDLLORT 7058
Qy 73 ----- 72
Db 7059 LSIDEQAPKIIKKLPEKIEPKEGEQAKLEVWVGPKPKVKWLRODEQIFASEEYQIENF 7118

Qy 73 -----LTSSLSV----- 80
Db 7119 EDGTSVLNVHYVDDLGTSPEAYNPLGVAVTTALFAVEGIVGSKOYRKPEWSQMEEM 7178
Qy 81 ----- 80
Db 7179 QVALKAACSPSLNEMRDCRAALGETAKFSIQFAGNPIDIQWYFNNVQLRASEKYRW 7238
Qy 81 -----CGW----- 83
Db 7239 VQEQTALIMKITSEDGCGYNCKLINEIGMTMTAKFDISSTSTIVSETKAKTTVKKKS 7298
Qy 84 ----- 83
Db 7299 GKKTWKRSGASESONVQKTEIRIIPTSAVETSMNVKVPVSLVEKSEISEVLVVKD 7358
Qy 84 ----- 83
Db 7359 REVADAERSQOLIEIEIEEIEEKVQHDEDEVEVQEQETYTSKKIEITKTVELI 7418
Qy 84 ----- 83
Db 7419 RTKISEKIITIEDVQVLSHHEEVQWLLSEIAESFGIGESALRDLATIGLLRYGCCHY 7478
Qy 84 ----- 83
Db 7479 EITYMYEQNIFISKPKPESQALVOLVEREGHEELISQILSESSNEDETILAATVGFKPF 7538
Qy 84 ----- 83
Db 7539 IRMIQYETIIEIVIRKVFREDFISQDWKICCKERIVETSQIIIESHEAITHVKIETATTK 7598
Qy 84 ----- 83
Db 7599 VEKLPKQEOHVQNLQEOEQVKIQVQTKQIAQMTKIKKHKKHQEQEVSETTIQCEQ 7658
Qy 84 ----- 83
Db 7659 KETLAHETSAELPQSETLEQIESLSYETLPIQNLKDTLTQTVAVSVTTELSTPSPTAS 7718
Qy 84 ----- 83
Db 7719 RVQEEILPQKVLAINEEVLPLDEFGLRKESPRKPKENKLTENIEVLKHALNVSHAKTAES 7778
Qy 84 ----- 83
Db 7779 SKELPSKIPKSVKQORKMKESRSLVVEAPNABEEAIEDLKPKAVSQEVQSDILFSHITE 7838
Qy 84 -----SQT----- 87
Db 7839 EQHQALETIEKLKPTSAIEDTVQOKLLSQEELIAEVLPSSETVGRDVTDRPPGETISPR 7898
Qy 88 -----NPED----- 91
Db 7899 LTPNMLCITEQPEDSIGEMQQAAKERMETPSMSVTSKAVGGOEVLNVDMHPLIT 7958
Qy 92 ----- 91
Db 7959 OPTKGLADYTIKAEVPVQVEIITFDSLRETVAKTQTAKSNALELPELSEGLVSSTAD 8018
Qy 92 ----- 91
Db 8019 SHSPIAEDLPIFEKDVKEATIDMQHQHVTTSSETVSNENAVKDKAVDTPKMAEGTLGQS 8078
Qy 92 ----- 91
Db 8079 SALTIGETQOMNLVETTVELIEPNVESTKPAKGALTEAYGTAESNEETLLESGLVPODN 8138
Qy 92 ----- 91
Db 8139 RKIEGQKNI SEGEYVAKVQTTVTDTTEGEFVSVA PKLVNPKFDFVQSALQIKODTTVE 8198
Qy 92 ----- 91

Db 8199 KEEILSSNIELAPOLATSNMPPAELKVTSIYEVOPGLTSSDIIITEOTKSVSANQVFETMS 8258
Qy 92 ----- 91
Db 8259 IGVTSKPDMLESTSHIDAFQHPFKTGTDLDENQOQPLEVTNVQITESSDIIIDVLPNQK 8318
Qy 92 -----DTDP----- 95
Db 8319 LTKAETVTDGPKYAEGLVWLPMESTIDKTEDTKPTAVNADISMHQFQFQTDVREOEPLEST 8378
Qy 96 ----- 95
Db 8379 LTRTEDLKPOQTTSQFGLLSLETSCVLTLEGESVLSVKERHPEOSAAITGSALQVA 8438
Qy 96 ----- 95
Db 8439 NITRPQHMESLDRLEQKVPYYQANVNIGETLPNVKIDSPDVLSDLNTPDYNKSKGR 8498
Qy 96 ----- 95
Db 8499 VOLJESTTSLKTTTAVVSESTEELKDLNITOPVHIKPKPYESDQKISIEQTNVLEHVS 8558
Qy 96 ----- 95
Db 8559 LNPVPALETIQSSIKSLHEINVRETIDILEKEESLKDVHDHISGLAKIILDCTTGIAQVR 8618
Qy 96 ----- 95
Db 8619 QEETLEHEEDLKAPLIPLEKAI PASSELHRLPLTEYVQEQGTSMDTDFKVSNNKASPNI 8678
Qy 96 ----- 95
Db 8679 DHLVETKSSSEMIYVDSSINSVDSEFPAGIVPKSLVFPFRHTMTVTVAFNASENFEILSA 8738
Qy 96 ----- 95
Db 8739 DQIATNVQDSLSQSI IAEQDIAFETEQNLGLETTPTHKPKLLKDDQNLHAKLVDEATVY 8798
Qy 96 ----- 95
Db 8799 EMQEQEKVKYNIQQAIEITHDLPOVYATDLOQTFEAEKEITTREQSYVAATTDIISRL 8858
Qy 96 ----- 95
Db 8859 GLAMTTKTHPVEGIDVLLSSPPKPSLAQTNVEETQHEVRVRETQAI SESEBELTDGRLLPV 8918
Qy 96 ----- 95
Db 8919 SAVESIDSTFKVTSDSQPPVFDKELSIPTVSPLEARAKPSLNLOGTTTTQVIPLESSV 8978
Qy 96 ----- 95
Db 8979 LLKQTHVAKAQOEYVAQVESNKVHVQMONLVNKHEDI FENABIEFNFKPITEGTOLET 9038
Qy 96 ----- 95
Db 9039 VVIEWPIDNVGGIHLAPQPSLTLATLTSTDIVNQSHVIDTQVPLEMESEAAQPLDNIAQ 9098
Qy 96 ----- 95
Db 9099 ARIKSAEDHVHTNVSEDTAQAAADELQSLVTEEVSVSSIOETVELKIPLOKTANLTQOT 9158
Qy 96 -----GHAD--- 99
Db 9159 PQNSVNVCOOLAYEETPDIAFEPHALTRATTSSVPTFLKPAENATVNIYENIEHGDKPK 9218
Qy 100 ----- 99
Db 9219 GTVNLTNSNLSNLSLVVSVQEVTSVPFSLGSLATVEPOELKAMPVTKSSTNLAYSEEKVG 9278
Qy 100 ----- 99

Db 9279 NQOEVTKIETVEEDDKQPETTVTVEELPFQOEKPEEIQEIPBEVRVTVTETDGPXXXX 9338
Qy 100 ----- 99
Db 9339 IRTRVIKKVGKQOEVTKIETVEEDDKQPETTVTVEEVPYEEBKPEEIQELPEEVRVET 9398
Qy 100 ----- 99
Db 9399 VTEDGPKKKIRTRVKKVKGDKQEVTKIETVEEDDKQPETTVTVEEVPYEEKLEEQ 9458
Qy 100 ----- 99
Db 9459 ELPBEVRVTVTETDGPKKKKIRTRVKKVKGDKQEVTKIETVEEDDKKAETTVTVEET 9518
Qy 100 ----- 99
Db 9519 ELSAPSGVKQLKRVIVQKPEDAVTFELPERKSVILSEKEDGTPTKTVIKRIKKIQ 9578
Qy 100 ----- 99
Db 9579 GPNNEVTKVQTVEEYKAPOTIVSVEKFNTPFPPELPEERLSEVVMPLPDEVFSEAVDEG 9638
Qy 100 -----L 100
Db 9639 RLKMKTKKRIIRKPALDNTVEEVEIGIIEQDNVEPIYSVKIQERPLTESKPEDSKLIEL 9698
Qy 101 VLYITRFDLELPGD----- 114
Db 9699 PEHVTELNVILPDGKKRRTVKSRAPKSLDDDLDEVTTIHIIEEDKEPLTKVNIIEVVP 9758
Qy 115 ----- 114
Db 9759 SDEISITPIEELPEETVTEELDENKPKKTKTRTFKKGPPDDDEYFOQTIDEEG 9818
Qy 115 ----- 114
Db 9819 KEPISLRVSDENIADIIDISKLDDKVLKHQKPHKHQDYKEYTITEPEEASADAL 9878
Qy 115 ----- 114
Db 9879 QKPTKDKTPKOKTLEPIEVEDETVIEDGTGEQTDQIAIKRPRKVQGNVQVEAVDEK 9938
Qy 115 ----- 114
Db 9939 PIEKKEKAKKKVVKTRDEMDDYIHFLLHQBEPKTVLPQVQRTMELPQARRDSSFQ 9998
Qy 115 ----- 114
Db 9999 PVKLTPMKIEKVEFKPKQVVEISSVVEFPQMLKLKAPKQRPQEBKKKKNEASFKNKKLKS 10058
Qy 115 -----NRQV----- 118
Db 10059 WIRFPVAPYCFYVVTLETNREVGLSRNVDEAEVLKLRPKFKHSHKPEKAELEAD 10118
Qy 119 -----RG----- 120
Db 10119 LGAYESDHSKNEKLLHPKYKRGKKEKIEPTDESRLKFKGKVPQNEEASEEVLNPKV 10178
Qy 121 ----- 120
Db 10179 KLDIAEIDAEMVPTQEEVVKKKPKKSSKPEEGLQFEPFPEEMERTSDIREESDTS 10238
Qy 121 ----- 120
Db 10239 VSDVTSTQEKPIYKKKKVTPSPQKNQYKILFGQPREFETPEDDLNRKQGERPDDDK 10298
Qy 121 ----- 120
Db 10299 ADTKLPKPFYFVLDEPGVQAEVPLSETIAKEPKKRIKVKTEQEDNTIEIVPLSP 10358
Qy 121 ----- 120
Db 10359 EDNDEQIFEITVTSSEIPOGDAKAKTIGKKVKRMKMOELDDFVTELOEPEPQEVYETRM 10418

Qy 121 ----- 120
Db 10419 SDFVEVKLTLPSEMSDOKPTKRLRHEKGEVQVLEIVSWAPGEEPFYEINVISSAN 10478
Qy 121 -----VTOL----- 124
Db 10479 TEGDSEBITTDKIKKSRKIKKDDLDAYIQOLINAEIPVTELEKYEKIDVDGKAKPKKL 10538
Qy 125 ----- 124
Db 10539 KAKTKKPIIDEGETLQGVTEHEPTKLUKTKKPEKKQVIEKELAEHAEVPEYDEFLINK 10598
Qy 125 ----- 124
Db 10599 TESERPOEKRVEETEKDIVPIVDKVLADNLCLPFVVVEEDLKMPLATDVIALEDEKII 10658
Qy 125 ----- 124
Db 10659 RKRVRKAKSKQVEIEIETEKFGDIPDEARVIVITTEVSGDTIDGPAPSTTEAPKXSV 10718
Qy 125 ----- 124
Db 10719 RKVKKEKLUKEFIVNIVEEAPLDHVEIYEDVLRTPPRESSEKEDIPSTTTTVEDEIVNP 10778
Qy 125 ----- 124
Db 10779 VLPEKIKTVDDVRVPKDKKKKIDNOKKIKISEFEPTPTSEDSTIEYTPKLSEHDEDLQT 10838
Qy 125 ----- 124
Db 10839 DEYSVDVKSLPKSKKSTKKOKESLPGPISLYTIRIETTPPEITEKIYEDGKEVVRV 10898
Qy 125 ----- 124
Db 10899 INKRRIKKAGPKPEYLIIEVIETYEDNNPEADVVRTIETTPSIDSKPOEDHKIQVVOEKK 10958
Qy 125 ----- 124
Db 10959 PKTESLDNVIQKLIQDOEIPQVDHKEFKATVLETSPESKAKKIKKHHKKTTEVIDGIPIT 11018
Qy 125 ----- 124
Db 11019 VIEVTIQTETDDEDFKPDVTLKEIDHENASEAPKVLKSKVSEKPKSKKSELEFKIA 11078
Qy 125 ----- 124
Db 11079 EEDKPKPVLEDISEDVQVQIIEEDGTQKQVEIKKKVSRKHGPKQEVFEITETKAIDEP 11138
Qy 125 ----- 124
Db 11139 LSEVTVEITDEQPOBEVLPAQEKKPIKKQKLPEDVNTYVVKVLEELTEPTQFETIPE 11198
Qy 125 ----- 124
Db 11199 DADKPKQPVIEDISENVQVQIIEEDGTQKQVEIKKKVSPKHGPKQEVFEITETRPSDE 11258
Qy 125 ----- 124
Db 11259 PLAETVITELTEGLNKDIVIPOEKKTVKKPKKLPEDIQSVIRVLEEFNEPQWAPASTE 11318
Qy 125 ----- 124
Db 11319 KPIIEDIAESIEIVPTEEDGITKEVEVKKKVSRKOGTKNQVFEIETKTSDEPLAEVT 11378
Qy 125 ----- 124
Db 11379 ILESGDKSOEVTILPKEKKPIKKTIKLPEDVESYVNVVLEEFCEPQSFESPEPTGEA 11438
Qy 125 ----- 124
Db 11439 HETKTKTKPKKPIVKAPENVILIEEMAPETVNIENVEIGSEVQVKTTLKKLKKEGPK 11498

Qy	125	----	124
Db	11499	LYLIEIKETYEENKEGDIETTTELVEPGSPDASDDQPVWQWIKKKPKVKDDLOKYI	11558
Qy	125	----	124
Db	11559	QOLIEQIITKPLBEYEBTEMDSKKPKVKSHNKKTIEVIDGLPVTIHEFNVEDIVSE	11618
Qy	125	----	124
Db	11619	PEDMETPKTLLEIKELPOLPDDSSKYLWNISDEFGADKPIKOPTQDQPIKKEKPLKKK	11678
Qy	125	----	124
Db	11679	KDVEYVPSLEAFDHTVKVWSEPTLEGTVEVTVKRKVSRKSGKDHIFEITETTSERD	11738
Qy	125	----	124
Db	11739	TAEVTVVELSSDEVLDSEKPKHERKIVKPKQLKODVEEYIINIIEEPIQIPVGLVE	11798
Qy	125	----	124
Db	11799	DEVEKQKEETKPKKSPITVIATEQEDNNDYDALVKEDLDQPIERALEKPPSPLEYTI	11858
Qy	125	----	124
Db	11859	SVEEDSVGEEOQPKPKKISKPKSIKQPSVDKSPDYLNVNISEESIIDEPIPEDVWTEA	11918
Qy	125	----	124
Db	11919	AEKPSSEPTPKVEBELEATEAVEKEVTDDDKGETTKQSVTRKIKKLVGPKBEEIIEIVETK	11978
Qy	125	----	124
Db	11979	TGDTPEYEVIVTTEEVQKSKEAPEKKAIVRKAKKIPKDDLDQYIOKLIIEQDIPTKEL	12038
Qy	125	----	124
Db	12039	EKYEKIDLDEPVKMKRKPIKKVKQSEEQPKSETEEPIEDKPVKEIseyseVSDSEPKLT	12098
Qy	125	----	124
Db	12099	AVKEPIEKPBEKPEIIVLEETVESKREPDBEGKREKVKVTKIKONRGSEVVHDI	12158
Qy	125	----	124
Db	12159	EEIDDTNESVITVTTPETPDQDQPSVKOKRTKKIKKDEVEDFVKRVIEEAPQEG	12218
Qy	125	----	124
Db	12219	SVDLVVIEDFVPKPSSEKRRKKPIKDKHTSVBEETPHEDEVLLIESVPDPSLSDLLTV	12278
Qy	125	----	124
Db	12279	VDSVPIEENKVNQIEDTKPEKKKKPKPSAKILEENVPEDTVEKPLEALHTSDLEK	12338
Qy	125	----	124
Db	12339	PVQFSISIKEEQKHTPEKKSSKISSEQPKOPSTEQYIISVTEHDLKPEEKPTV	12398
Qy	125	----	124
Db	12399	QVIQSETNVEETKDDTGKVKHQTTRKMLRRPAGEGEIIEIIVVRDDQPEAEITIVIEYEP	12458
Qy	125	----	124
Db	12459	EPVNODEKPKPKKTRVKKDDIHDYIOKLIIELETPKTELEKYEKIEFEPVVKOKPLDS	12518
Qy	125	----	124
Db	12519	PIDVLDESPEVOKKDKKRTICKGOCVEEBAPEQEIIPVQILEVVPVEDVKEVITEDG	12578
Qy	125	----	124

Db	12579	KPVOEKTTRKVLKKIGPERQTTFKITMIESEONDSTVIVUDEEPEIASPOSIEEHPEQSK	12638
Qy	125	----	124
Db	12639	EKLAPKPKTVRKVKKDDLSDYVVKLJEBEIPKVDLEKYEKVEKPKVLTVSDSIPEE	12698
Qy	125	----	124
Db	12699	PKDKSQPISVLPDPTTKPKKTKTPKPTKTEDTQQVPDEPTTVDTPDIPELTPTQTAQ	12758
Qy	125	----	124
Db	12759	PEDTATAQITPSAOEKSSTODDTKOTIQTVKHKTKTPDTOKSVETSSELPEVHKDYOISI	12818
Qy	125	----	124
Db	12819	IHEELVEEEOPEKILEVRVIDEVAEVEESQPIVEEVEDEEPQATEETVEDVTPKSKKK	12878
Qy	125	----	124
Db	12879	KVKKKTDDHDELIKKMLEOEIEKTELEKYEKIEFDVPKKLKPEFAALEPIKIERKEQKP	12938
Qy	125	----	124
Db	12939	TKVTILDATDVPKTVKLPKSKRKEKPAELTVOLPKFRLKARMLVEYPPAPLIPKTTDI	12998
Qy	125	----	124
Db	12999	GAIKDNGELSRNIGEABEILKPKPKTKKIKKIDDLKLEKYEKYEKYSSEBEPEBETP	13058
Qy	125	----	124
Db	13059	YKPEKAPKEEQEDVKLKLKGKKKPBEEAPENVTLKNIPOKPOVEEVEELKOKPK	13118
Qy	125	----	124
Db	13119	EVEIVEEQTKPKDGEFVPEFSEFDRPEYVPDELEQIEHPEIPEKVKPKSKTKYKPK	13178
Qy	125	----	124
Db	13179	DKSKSEPTIVSEIVAGVPKEEBAIPEQDVKFRKPERDAPETDSEIKLRVPQASKDEN	13238
Qy	125	----	124
Db	13239	PDEQALVTPKAEEPIPOEIEDKAIDDEKPKKSKPKVQPKQEOBIAKEEPEEVEVKEE	13298
Qy	125	----	124
Db	13299	EALVDKPIEIEKPKDKVKKEKKKPEAPVSEVVVIEBEPKPEEVPPEIIPVEKITTTLVLEP	13358
Qy	125	----	124
Db	13359	EDAPKEHQVKVIDPDERQETTEEVEIEKVTTRKKKPKQOPEBFEVTLKEPKEBQIOPDV	13418
Qy	125	----	124
Db	13419	VSAEISLPIEBEPEKQPEYVELKITOTTPPEPNDOIVAVKEKVKTRPKVKKEDKIVV	13478
Qy	125	----	124
Db	13479	EAEEEKQPVETIVEVEKQBEKKKSEKPKSYEFKISQTSIEEKPIEVAEBAPEETPKV	13538
Qy	125	----	124
Db	13539	EKKVAEKFSYEFTLKETDBEKVITVDDQPEEAPVSEVVVFKKPKPEAVEAEFVMTBPK	13598
Qy	125	----	124
Db	13599	IVEETSVETAIKOKTKKKDEBEAQLAIKWVESEAPVABEVFSEAPESKIVBEEVIAE	13658
Qy	125	----	124

Db	13659	EKPKEFTIRVSESEPKPEEPSVBOFTVYKKRKPSTVTFADEPATEIVIKESKPAPVWTTEDAH	13718
Qy	125	-----	124
Db	13719	IKTKKKKKYDVEABELKIKITEEVPQEIPILEEUSEEEVITETKKTAPVVEEYTKIG	13778
Qy	125	-----	124
Db	13779	IKETEPEKPAEAVIEEEEPVVTEPIEBAPKEPVFEHKRVIEETPRELVEEVIEBEKV	13838
Qy	125	-----	124
Db	13839	IRKKKPKPEIKEEPEAEVTVSTPKPVEEVEATSIAVIPQPTEEEAADLKITIEETP	13898
Qy	125	-----	124
Db	13899	POELVQIEIEIEVEEPKABEQPTDFTFATKQSEKKPTVEELPEEQVTIQKKKKAPVP	13958
Qy	125	-----	124
Db	13959	EVVEEPAEFLVKPKTPVQEVTEBAKITKSKKPKVKEBAAAELKVITIEPIPEVQEI	14018
Qy	125	-----	124
Db	14019	IEEIEIEEEKPAEVVIEVKESQPEAVEDKEVSLPKKKPKAPIVEEPEABITLKPVKSE	14078
Qy	125	-----	124
Db	14079	EVQBEAKIVKKKKPIDEVADELTVKVEEVVPEPIVEEVIEEPIEIKKKPEPEPD	14138
Qy	125	-----	124
Db	14139	IVDAAIVKLKXPEPDADEVVAEVLTKPKAKTEVTEBEFSVDVKLPKEKKERPVEIEEE	14198
Qy	125	-----	124
Db	14199	IIIEAVVIRKKPKPFPEPTVEDLEETFSLSFKXPHTINGEVBAAVLVKRPVKVPTLD	14258
Qy	125	-----	124
Db	14259	EAAAELSIKROEEYECEGDIIEFVVSQORPKPLOITEEDEEAYTVKCLKRRKQVDIPE	14318
Qy	125	-----	124
Db	14319	YADVENVTFRARSTKTKEDVDQEFNIALDSYAEBSISMGVKLKKPKIKTKTFSEAADEAK	14378
Qy	125	-----	124
Db	14379	IKIIODFDDGSEPIIEEIRDDEDTIDEVEEPEEYFVEELPPDEVDFKLPKKHKPKPAYSV	14438
Qy	125	-----	124
Db	14439	QDEEEEQFLIGIRHPKRDSTVYDEDSLTKFKKRVVQOLFNEESENLMYSICNYADN	14498
Qy	125	-----	124
Db	14499	NEAINLVEGEKVTVVGRHSSEWVYVKSTTEEGWVPAOYLMEPEEYAYQVQNKLEKID	14558
Qy	125	-----	124
Db	14559	KLPVFERGPKDPIAPRIEKLQPIHTPDGYTVQFECKVEGNRPQIAWFRETAIIKPS	14618
Qy	125	-----	124
Db	14619	QDFQFYDDDNVATLIIREVPFEDAGQFTVVAKNAGFTSTTELIIVESPLSDHGSDATA	14678
Qy	125	-----	124
Db	14679	LSRRSMSRESSLADILEGIPPTFSKKPKAQYVDENTNVILECRLVAVPEPDIVMTFNGED	14738
Qy	125	-----	124
Db	14739	IDEEIEKNVRIVTESDMHMYCSVWHISKVKSQEGTYEVIAITNREGEARLPITLKVRTTD	14798

Qy	125	-----	124
Db	14799	KEAPQILEPLRNWVIREGESVILSTQIVGNPPKVTWYKDGKPVKNKAKSDKDLHTLTIT	14858
Qy	125	-----GCACSP-----	130
Db	14859	POKSEGEYTKAVNPLGSVETTANULTIEBPAGNAEPLFVFPBEEQNVPOKGEIRLPA	14918
Qy	131	-----	130
Db	14919	KVSGNPVEQWLFNNLTPLFPPSERIOOVYDENIELIKDANPETDGDYKCIASNPICK	14978
Qy	131	-----	130
Db	14979	TSHGARVIVEDEVTFTKKLKXITIEBVOSLTLECTSHVVTWKFFNGKELSGMDHRV	15038
Qy	131	-----	130
Db	15039	VVEDGKTHKLVRNTNLRDSTGYCKVKQETQSTVEVLQRPDKVLEDEYVEKDTA	15098
Qy	131	-----	130
Db	15099	ILDVELTTEATEVTWYKXGKITPENKNVEFIKDGKARRLVIRDVTHDEGOYTCKIEGQ	15158
Qy	131	-----	130
Db	15159	ECSCELVIELPPEIPEPLNDVAVTKGENAVFEVLESGDALVKWFKNGKEIVFNERIQL	15218
Qy	131	-----	130
Db	15219	AIDGKKQSLRIVKAKPEDVGEYSVOVGEQTSKAKLTVEBPLVDFVIRLPDITLATKTDA	15278
Qy	131	-----	130
Db	15279	ECTVOLSQDVEVTWCKKGDKPKNQKHEVFEVGTVRRLVIHDASDEDAGEISCAENV	15338
Qy	131	-----	130
Db	15339	SSTKLVCBELKLPVITSDKQTIKVKENDDDTVTKYGTGVTPEACWTTRKVIKSKR	15398
Qy	131	-----	130
Db	15399	TIPTIDEOSAKLTIKKVVDDDEGEYTVKLVNPGAEASLHLVIMRKPTAGTPOPLEIM	15458
Qy	131	-----	130
Db	15459	HDSITLYWKAPEDDGKSEIEVILEYQDVKEEKWTEIRKIDTYYTISKIDTEYVFRS	15518
Qy	131	-----TWSCILIT-----	130
Db	15519	IAVNEVGPPSPPLSPPIRLVPKVEKAPSVQEPLODVVSELDKBEVTLSCVFGGIPBKV	15578
Qy	131	-----TWSCILIT-----	137
Db	15579	TWKQNGQVFSRSIRYENRVAKYTIETKTIEATYTCVATNEKSAETSCRLKLOOKPV	15638
Qy	138	-----	137
Db	15639	LEVEDKYLTKLRTGSLTIPATVRGYPQPTVWHKETIEOKTKSVTIETTSTVTV	15698
Qy	138	-----	137
Db	15699	KKVTRQSGKYKVTATNESGTYVECTVOVDKPSRPOSLEIKDIKKDSIVLEWTPPVDD	15758
Qy	138	-----	137
Db	15759	GGLDIEKYTLEKCDVQNNVMKVSDFNKDIKSYAVQKLSMNAQMYRVVAANPIGESEPT	15818
Qy	138	-----EDTGDL-----	144
Db	15819	ESDPVITTKFKEPGPPRGPTTVTSGMNDTSFNLAWPSETDGGSKLIEYIYVEIRETETT	15878

```

QY 145 ----GVTI 148
DB 15879 YRSVGVTL 15886
|||:
RESULT 10
OBS40 PRELIMINARY; PRT: 6889 AA.
AC Q8XS40;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Probable peptidase synthetase protein.
GN RSP0641 OR RS05860.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangerot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gapin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502 (2002).
DR EMBL: AL646079; CAD17792.1;
DR InterPro: IPR001227; AC transferase.
DR InterPro: IPR000954; Aminotran_3.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR002103; Bac luciferase.
DR InterPro: IPR001242; Condensatn.
DR InterPro: IPR000276; GPCR Rhodpsn.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR003880; Ppanctn_attach.
DR Pfam: PF00698; Acyl_transf; 1.
DR Pfam: PF02022; aminotran_3; 1.
DR Pfam: PF00501; AMP-binding; 5.
DR Pfam: PF00296; bac luciferase; 1.
DR Pfam: PF00668; Condensation; 4.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF02801; ketoacyl-synt_C; 1.
DR Pfam: PF00550; pp-binding; 6.
DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00075; ACP DOMAIN; 6.
DR PROSITE: PS00455; AMP_BINDING; 5.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; UNKNOWN_1.
DR PROSITE: PS00012; PHOSPHOPANTHETINE; UNKNOWN_4.
DR PfamId: Complete proteome.
SQ SEQUENCE 6889 AA; 744917 MW; A7F31DCFFDED1D37 CRC64;

Query Match 30.9%; Score 239; DB 16; Length 6889;
Best Local Similarity 1.5%; Pred. No. 8.;
Matches 91; Conservative 27; Mismatches 29; Indels 6063; Gaps 27;

QY 1 AAGGILHLELLVAV----- 14
DB 328 AAYGMAEATLLIAFGWAIRFOGPRCLPFSRLALQKMAVAADDEADRTALASHGSLTC 387
|||: |||:
QY 15 -----GPDVFQ----- 29
DB 388 HQLAIVDPETLRRCPCRGVEIIVSGFSVAQGYWRRDEDSRRSFAGEMAEPADGERYLRT 447
|||: |||:
QY 30 -----VLTNLI----- 36
DB 448 GDLGFLHAGELYICGRKLDLIILNGLNIYPQDVLAFAFESHARLRENGTIAFAVDRDDTE 507
|||: |||:

```


Db	1588	SVEPSRAVTSMPVAAAMSSSVNSASRAQAIQOALRCMVAKMLHADVARVDPDLPLELGL	1647
Qy	78	-----	77
Db	1648	ADSLMMVQAIGSTEQTYGTVTVRQLFEELTTIAADYVDRQMPREARLEAAEAPATV	1707
Qy	78	-----	77
Db	1708	PAAAVMPMPAPAPAPFAIAVDARPDVPLAAIPALPOSSLERLLSQOLDALSOLTAR	1767
Qy	78	-----	77
Db	1768	QLALFOGGTVAATAGGPASVAGETSPVPTVMPAAVPPVAPAPASTAAFGGTHKPPVPQ	1827
Qy	78	-----	77
Db	1828	PVRTAKADPFRDLSARQRAYLDGFFVARTERTRGSKALVQRYRPVLADNRVSAGFRFST	1887
Qy	78	-----	77
Db	1888	KEMLYPVVSRSEGYLWDPDGNVYIDLTMFGVGNLFGRHPAFVQEAALDAQLKTGLELGP	1947
Qy	78	-----LSVC-----GWS--84	84
Db	1948	OTRAGEVAELVTRLTGLERVAFVNCSTGTEAIMLALRLARTVTRDKIVVFAGSYHGSDD	2007
Qy	85	-----	84
Db	2008	TLMVADGACTVAMAPGLQPCASAHITVLDYGAPESLARIHHAHELAAVLVEPVQSRPH	2067
Qy	85	-----	84
Db	2068	WQPREFLHALRELTREHGIALIFDEIIMGFRLHPGGAQAWYGEADWATYKVGIGGMPA	2127
Qy	85	-----	84
Db	2128	GMVAGSAAYLDAVDGGQWRVGDASYPOADTTFYAGTFCKHPLMLVAARAVLQRLAQGPA	2187
Qy	85	-----OTIN-----88	88
Db	2188	LOETLNARTAEVRLNGVFADARVPVRVAVHCGSSFRLEASPSIDLLYYHLLAGGLIYW	2247
Qy	89	-----	88
Db	2248	EGRGMFLSTAHDADVDVRVVEIFADSVRAMLDGFFEDGAPTTPSGGGGRSPFAARVGG	2307
Qy	89	-----	88
Db	2308	PASTDAPVAPALAEAAVARLSAAPPATGGIRFGISFFGHYASGYDAQKYRLLFEAR	2367
Qy	89	-----	88
Db	2368	YADAGFSSLWPERHFHAFGLSPNPSVLSAALARETSHIQLRAGSVVPLHHPVRVAE	2427
Qy	89	-----	88
Db	2428	EWSMVNDLSOGRVGIACASGHHPNDFVFAPEAFGSHRELMPQRIEQIQLWRGEPLVRD	2487
Qy	89	-----	88
Db	2488	GSSREIEVKLFPMPROPELPITWITVGNPDTYRRAGEIGAGILTNLMQTVAEALCNLAL	2547
Qy	89	-----	88
Db	2548	YQALVEHGHGVERSRSVLLHTFVCDEAAEARAVARAFFIHYLRSSVGLFQNMVDSLGL	2607
Qy	89	-----PEDD-----92	92
Db	2608	QADVSTLSEDDRDYLLSVAYERYEHSALIGSPATCRALVERLOAIGVDEIGCFIDFGVD	2667
Qy	93	-----	92
Db	2668	PDTVLRDLQDALLKQSFETAAVDDDAERYPLVPAQKGIWFECQISHEAALSNTTNTVL	2727
Qy	93	-----	92
Db	2728	GLRGALDHAAALARALQQVVDHRHAALRSVVEADGEHQRVLPVEVGLPLVDFSRADRDAA	2787
Qy	93	-----TDPGHADLV-----101	101
Db	2788	IGQWFDVNNHRPMDPGHGPLVRACLLRKGEAEHLAITFHHVIIDGYSQEIVLOELAAQY	2847
Qy	102	-----	101
Db	2848	RAACRGAGPLPAASPFEQVERHQAYLRSRYOODRAYWRGQFGLPPSLELSGRHARP	2907
Qy	102	-----	101
Db	2908	SAPSHRRARRHLLTIDGERYARLQQLSRKLGGLPMTMLAGVAVLLQRLSGOQMVIGVPM	2967
Qy	102	-----	101
Db	2968	VVGRAGSEASLVGCTLNLVVPVRCDSGSPFAEFLGRIKRSVLEAHAAHADYPFGHLLRD	3027
Qy	102	-----	101
Db	3028	LDLRASORRPLAPVLFNLNLSLALPQFDALQAWLEQSPISFPDDLTTIDVMQLPDLQVM	3087
Qy	102	-----	101
Db	3088	FOYQEVLFEEHAEIERMAAQFVQLLDGIVADPACSTGRLLALLSABERROIVDMVNAGEPVP	3147
Qy	102	-----	101
Db	3148	EPSETLHAFVFAQVRRTPETAIVEHEGWRSYAEELNARANRVAHALIGLVGPDARVGLC	3207
Qy	102	-----	101
Db	3208	AERSVELVIGLLGILKAGGVVPLDPSYPRDLAYWLEDSAPVAVLAOSNTREPLGALSV	3267
Qy	102	-----	101
Db	3268	PVLDLENPLEGEAEHDPQVTGLEPNHLAYVIYTSGSTGPKGVLEHROVARLFTSTOPW	3327
Qy	102	-----	101
Db	3328	FGGAEDVWTLPHSFADFVSMELFGALLHGGRLVVVPKLTARSPQAFYALLCEAGVTVL	3387
Qy	102	-----	101
Db	3388	NOTPSAFQRLMAAQOEAPAAHRHLRQVILGGEALVGLRPMWYERAEHAGTOLANMYGIT	3447
Qy	102	-----LYT-----104	104
Db	3448	ETTVHVSYPALAEADAQGTGSPIGRRIPDLRVVVLDAHGEPPVGVGTGEMYGAGVARG	3507
Qy	105	-----	104
Db	3508	YLNRPETLTAERFVWNPFGHEGRMYRTGDLGRWLPDGSLEYQGRADAQVKLRGFIELG	3567
Qy	105	-----	104
Db	3568	EIBASLSQCAGVREAVTVTVREDVPGEORLVVYVVSGEAIEAQNALRELOGSLPAYMVPA	3627
Qy	105	-----	104
Db	3628	YVRLAHLPLTNSKGLDRKGLPAPEGHAYASTAYEAPQGEVEQTLAGIWOTLLGVERVGRH	3687
Qy	105	-----	104
Db	3688	DDFFALGCHSLQAVRLVTVVOLGAEGLTALFAQPSLSAVAQAIVRGGSALQAITAA	3747
Qy	105	-----	104
Db	3748	DRSEALPLSPAQRRLWLLAQMEGGSAYHIPVGLRLKGLDELDALGRALDRIVARHEALR	3807

Qy	105	-----TRFDE-----	110
Db	3808	TRFEVREGAIFORVASADVGFALDRVDLQCGADREQTAAALSEREANTPPDLEQPLRG	3867
Qy	111	-----	110
Db	3868	CLVKLGEQEHVLLITMHHIVSDGWSQGVLAARELGALYEAIRSGGDDPLPALPIQVADYAV	3927
Qy	111	-----	110
Db	3928	WQRRWEGGELQRCAGYWEQALAGAPTLLSLPTDRAPAQDYAGGSVEVVFDETLSDL	3987
Qy	111	-----	110
Db	3988	KRLSORHGTTLFMTVLGWSALLSRLSGQEEVVVGVSPVANRTRSEVEGLIGFFVNTLALR	4047
Qy	111	-----	110
Db	4048	VNVGSATVSELLDRVKAKVLEAQAHQDLPFEQVVERVKPVRSLSHSPIFOAVFSWHNTEA	4107
Qy	111	-----	110
Db	4108	VDLSRLALSLEARENATAKLDIQLELAEDGRIVGTINATALFERSTAQRVADYLQR	4167
Qy	111	-----	110
Db	4168	MLQAMVADDGQOVGRIALLGEAERAQVLOAMNATERANWPAATLPALFEAQVWRTPDVAL	4227
Qy	111	-----	110
Db	4228	KHADQOVSRELDARANRLAHLRELGVAAADVGLVCVDRSIEMI VGLLILKAGGAYVP	4287
Qy	111	-----	110
Db	4288	LDPDYPQARLAYIFODAMLSVLVSKRALAQLPIANTQVVELDDAEPAWADYPPPPQVQ	4347
Qy	111	-----	110
Db	4348	GEFGQLAYIYITSGTGQPKGVAVTHQVASLVYSQCERFGVSSQSRVLQFASISFDAAV	4407
Qy	111	-----	110
Db	4408	SEIGHMLCGACLVLAQAQALMGAALTHLDRERITHVTLPAPVLAALMPEQALPADCHL	4467
Qy	111	-----	110
Db	4468	IVAGEACPPSLVRLWSEGRMTINAYGPTATVCATMSRALTADAPSIGGPIGNVRVYVL	4527
Qy	111	-----LPD	113
Db	4528	DAYLQVPVGVVTGELYIAGSLGARGYQWQAGLTAERFMANPFALGERMYRTGDLGRWLPD	4587
Qy	114	GN-----	115
Db	4588	GSLEYQGRADAQVKLGRFRIELGEIEARLSQCAGVSEAVTVREDAPGEORLVAYVVSGE	4647
Qy	116	-----	115
Db	4648	AIEAOTLREQQSRLPPEYMPVPAAYVRLEHLPLTPNGKLDKRGKLPAPGQAYASTAYEAQ	4707
Qy	116	-----RQVR-----GVTQL-----	124
Db	4708	GEVEQTLAGIWTLLGVERVGRHDDFFALGGHSLQAVRLVAQVRTQGLGELGLTELFAQP	4767
Qy	125	-----	124
Db	4768	SLSAVAQAIVRGQGSALPAITVADRSGLPLSFSQORLWFLAQMEGGSEAYHIPVGLRLK	4827
Qy	125	-----	124
Db	4828	GELDEDALRRSLDRIVARHEALRTRFVTEEGQAVQVRVASADVGFDLDCVDLQCADREOA	4887
Qy	125	-----	124
Db	4888	LATLSEREANTPPDLAHGPLIRGLRVKLGEQEHVLLITMHHIVSDGWSQGVLAARELSLY	4947
Qy	125	-----	124
Db	4948	EAYRAGNADPLPALPIQYADYAVWQRRWLEGGELQRCQTYWEQALAGAPTLLSLPTDRAR	5007
Qy	125	-----	124
Db	5008	PPQDYAGGSVEVVFDAELSTGLRTLSQRHGTTLFMTVLGWSALLSRLSGQEEVVVGVSP	5067
Qy	125	-----CGA-----	127
Db	5068	VANRTRSEVEGLIGFFVNTLALRVEVGATVSELLERVKAKVLEAQAHQDLPFEQVVERV	5127
Qy	128	-----	127
Db	5128	RPVRSLSHSPVQFQAALSNLWNTAEVGLSLELEGLTIEGVDAQAAKFDLTLELRETSEGL	5187
Qy	128	-----	127
Db	5188	AGSLDYATATLFDRTATERYLGYLHRLKAMAADDSQEVNRIALLDEGERTOLLESWNETK	5247
Qy	128	-----	127
Db	5248	APYPRASTIHGLFEAQVVRTPEAIVVHEGQVSYAELNARANRVAHALRRLGVGPDAV	5307
Qy	128	-----	127
Db	5308	GLCAERSVELVGLLGLTKAGGGYVPLDPSYPQDRLTMYLSDSAPVAVLTQSTREQLGA	5367
Qy	128	-----	127
Db	5368	LSVPVLDLGGLEBAEHDPQVTGLEPHHLAYIYITSGTGRPKGMNEHGRVGNRLMWAQ	5427
Qy	128	-----	127
Db	5428	QTYRLDASDRVLOKTPFGFDVSWWELFWPLLAGARLVNARPEGHDPAYLAATIEQAGIT	5487
Qy	128	-----CS-----	129
Db	5488	TLHFVPSMLQLFLDOVEAGRCQGLRRMLCSGEALSHALQQRSLARFPHSELHNLGYPTA	5547
Qy	130	-----	129
Db	5548	AIDVTAMRCNAEIHPPGVVPIGRPIANTQMYVLDGHQOPVPLGVTGEIYIGGIVARGYLN	5607
Qy	130	-----	129
Db	5608	RPELTAERFVYNPFHGEGRMYRTGDLGRWLPDGLSLAYQGRADAQVKLGRFRIELGEIE	5667
Qy	130	-----	129
Db	5668	ARLSQCAGVSEAVNAMEDVPGEORLVAYHVSDEAIEAQTURELOQASLPEYMPVPAAYVR	5727
Qy	130	-----	129
Db	5728	LEHLPLTPNGKLDKRGKLPAPDGOAYASAAEAPQGEVEQTLAGIWTLLGVERVGRHDDF	5787
Qy	130	-----	129
Db	5788	FALGGHSLQAVRLVAQVRTQGLGELTELFAQPSLSAVAQAIVRGQGTALPAITVADRG	5847
Qy	130	-----	129
Db	5848	EALPLSFAQORLWFLAQMEGGSEAYHIPVGLRLKGEDELDRRLSLDRIVARHEALRTRF	5907
Qy	130	-----	129
Db	5908	EYQEGQAVQVRVASADVGLTLDWADLSAESAASEHQLGLLAEARAPFDLEQGLIRGLV	5967
Qy	130	-----	129

Db	5968	KLGEQEHVLLITMHHIVSDGWSQGVLAELGALYEAYRAGNSDPLPALPIQVADYAVWVR	6027
Qy	130	-----PT-----	131
Db	6028	RWLEGELORQGTWYEQALAGAPTLISVPTDRARPPQDYAGGSVEVFDETLISAGLRKL	6087
Qy	132	-----WSCIIT-----	137
Db	6088	SORHGTTLFMTVLGNWSALLRSLSGQEEVVVGSPVANRTRSEVEGLIFPVNTLALRVEV	6147
Qy	138	-----	137
Db	6148	GSATVELLGRVKSRLVLEAQHODLPFEQVVERVPVRSLSHSPVFOAALSWLNTAMVL	6207
Qy	138	-----	137
Db	6208	KLGLTIEHLADESYTAKFDLSLALREHGRGRIAGSLDYATALFDRETIERYLGLHRLLA	6267
Qy	138	-----	137
Db	6268	AMVENDSQVSRICLLDEGERAQLESWNETKAAYPDASTIHGLFEAQRRTPEATAVEH	6327
Qy	138	-----	137
Db	6328	EQQVSYAELNARNRVAHALRLGVGPDPARVGLCAERSVELVVGILLKAGGVVPLD	6387
Qy	138	-----	137
Db	6388	PSYPQDLAYWLEDSAPVLAQSNTRQGLGALSVPLDLDRLLBEAEHDPQVTGLEPH	6447
Qy	138	-----EDTG-----	144
Db	6448	HLAYVIYTGSGTRPKGVWVEHGRSLNLDWYLEDGLRSDDTVLLVSSYSFDLTOKNIL	6507
Qy	145	-----GVT 147	
Db	6508	GPLLAGRLHLAGFPVDPDALVAQIRREGIT	6537
RESULT 11			
Qy	145	PRELIMINARY; PRT; 10223 AA.	
AC	Q54296		
DT	01-NOV-1996	(T-EMBLrel. 01, Created)	
DT	01-NOV-1996	(T-EMBLrel. 01, Last sequence update)	
DT	01-JUN-2002	(T-EMBLrel. 21, Last annotation update)	
DE	Polyketide synthase.		
GN	RAPB.		
OS	Streptomyces hygroscopicus.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
OX	NCBI_TaxID=1912;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NRRL 5491;		
RX	MEDLINE=95372374; PubMed=7644502;		
RA	Schwecke T., Aparicio J.F., Molnar I., Koenig A., Khaw L.E.,		
RA	Haydock S.F., Olinyk M., Caffrey P., Cortes J., Lester J.B.,		
RA	Boehm G.A., Staunton J., Leadlay P.F.;		
RT	"The biosynthetic gene cluster for the polyketide immunosuppressant		
RT	rapamycin.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 92:7839-7843(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NRRL 5491;		
RX	MEDLINE=96186895; PubMed=8635730;		
RA	Molnar I., Aparicio J.F., Haydock S.F., Khaw L.E., Schwecke T.,		
RA	Koenig A., Staunton J., Leadlay P.F.;		
RT	"Organisation of the biosynthetic gene cluster for rapamycin in		
RT	Streptomyces hygroscopicus: analysis of genes flanking the polyketide		
RT	synthase.";		
RL	Gene 169:1-7(1996).		

RP

SEQUENCE FROM N.A.

RC

STRAIN=NRRL 5491;

RX

MEDLINE=96186895; PubMed=8635730;

RA

Aparicio J.F., Molnar I., Schwecke T., Koenig A., Haydock S.F.,

RA

Khaw L.E., Staunton J., Leadlay P.F., Staunton J., Leadlay P.F.;

RT

"Organization of the biosynthetic gene cluster for rapamycin in

RT

Streptomyces hygroscopicus: analysis of the enzymatic domains in the

RT

modular polyketide synthase.";

RL

Gene 169:9-16(1996).

DR

EMBL; X86780; CAA60459.1; --

DR

HSSP; P25715; 1MLA.

DR

InterPro; IPR001227; AC transferase.

DR

InterPro; IPR02085; Adh_zn_family.

DR

InterPro; IPR004410; FabD.

DR

InterPro; IPR001899; Gram_pos_anchor.

DR

InterPro; IPR000794; Ketoacyl-synt.

DR

InterPro; IPR003880; Ppantne_attach.

DR

InterPro; IPR002364; OOR_zeta_crystal.

DR

InterPro; IPR000634; S/T_dehydratase.

DR

Pfam; PF00698; AgyL_transf; 6.

DR

Pfam; PF00107; adh_zinc; 1.

DR

Pfam; PF00109; ketoacyl-synt; 6.

DR

Pfam; PF02801; ketoacyl-synt_C; 6.

DR

Pfam; PF00550; pp-binding; 6.

DR

TIGRfam; TIGR00128; fabD; 6.

DR

PROSITE; PS00075; ACP_DOMAIN; 6.

DR

PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.

DR

PROSITE; PS00165; DEHYDRATASE_SER_THR; UNKNOWN_4.

DR

PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_4.

DR

PROSITE; PS00012; PHOSPHOPANTHETHEINE; 6.

DR

PROSITE; PS01162; OOR_ZETA_CRYSTAL; 1.

KW

Phosphopantetheine; Transferase.

SQ

SEQUENCE 10223 AA; 1072168 MW; 8631C61BDD4E16EC CRC64;

Query Match

30.8%; Score 238; DB 2; Length 10223;

Best Local Similarity

1.0%; Pred. No. 31;

Matches 104; Conservative 17; Mismatches 25; Indels 9871; Gaps 32;

Qy

1

AAG-----

3

Db

86

AAGKSYCVGGFLDSAGGFDAFFGISPREALAMPQORLVLEASWEAFERAGIEPGSLR

145

Qy

4

3

Db

146

GSDTGVMFCAYPGGYGVGNADLGFGNTAGAVSLGRVSYFFGLEGPVTVDTACSSSLV

205

Qy

4

3

Db

206

ALHQAAYALRQGECSLALVGGVTVMATPQSFVEFSRQRGLASDGRCKAFMPSADGTGWAE

265

Qy

4

--GILHLELL-----VAV-----GP-----

16

Db

266

GAGVLLVERLSDAQRKGHQVLAVRGSAVNQDGASNGLSAPNGPSOORVIRAAALSNAGLS

325

Qy

17

----DVFOAH-----

22

Db

326

TAEVDVVEAHGTTGLDPIEAQALLATYQDREOPLLLGSKVSNLGHQTAAAGVSGVIK

385

Qy

23

22

Db

386

MYMALQHLVLPRTLHVDEPSRHRVDTGDAVALVTENQPMQPRRRAGVSSFGISGTNA

445

Qy

23

22

Db

446

HVILESAPPTQAVDDVPPEAPVASELVLVLSARTLPALVEYEDRLRAYLAASPGVDV

505

Qy

23

22

Db

506

RGVASTLAVTRSVFEHRAVLLGGDTVTGTTVSDPRVVFVPPGQSGORAGNGEELAAFPV

565

Qy

23

22

Db 566 FARIHQVMGLDVPDLEVNETHYAQPALFALQVALFGLLESWGVPRDPAVVGHSVGLAA 625
Qy 23 -----QEDTERVVL----- 31
Db 626 GYVSGLSLEDACTLVLSARARLMQALPPGGVMVAVPVSEDEARAVLGEVIEIAAVNGPSS 685
Qy 32 ----- 31
Db 686 VVLSGDETAIVLQAAAALCKSTRLATSHAFHSARMEMLEEFRTVAERLTYQTPLRMAAG 745
Qy 32 ----- 31
Db 746 DRVTTAEYVVRQVDTVRFGEQVASYEDAVFIELGADRSLARLVGVAMLHTDHEAQAAI 805
Qy 32 ----- 31
Db 806 SALAHLYYNGVTVDWNTALLGDAPATRVLDLPTYAFQHORYWLEGADRAAAAGHPLLGPV 865
Qy 32 ----- 31
Db 866 GLAEASGVLTTSQVSRSGDLWLQDOTVLPATAFEMALAAADETCGGLVEELTVEALLLL 925
Qy 32 ----- 31
Db 926 PDGAVQVQTVWSEPPDAGRRQLSIHARYSDDELWTRLATATLATTGTVSGWQAGEAWP 985
Qy 32 ----- 31
Db 986 PAGAVPEARVPSLRGVWRRGSEVFAEVALDDTHDTRFLHPGLMNTAAFAAVGEETPAA 1045
Qy 32 ----- 31
Db 1046 WOGLTLHARNAELRVRLTSQDDETLSEAVDSTGLPVLTAESLALDAPVNEPATSDT 1105
Qy 32 -----TNLNIGA----- 38
Db 1106 LUTLTWAGIATRQQTGLTVGAFEDLAADGAPVDPVVVFTALPGNDPDLAQTRILTAQV 1165
Qy 39 ----- 38
Db 1166 LRTVOEWIGGERFSDTLVVRTGTGLAAAVSGLMRSAQSEHPGRFILVESDDALTPDQ 1225
Qy 39 ----- 38
Db 1226 LAATVGLDEPLRLISDGRYEVRLTRAHADEPEPERAWPDGTVLITGSSGVLAGIVARH 1285
Qy 39 ----- 38
Db 1286 LVAERGVHRLLLSRCTPDRALLSELAEELGAAVDTAAACDVSDRAELARVLARVSPEHPLT 1345
Qy 39 ----- 38
Db 1346 AVIHTAGVVDDGWVESLSAQRLTETFRPKADGANHLHELTRDADLAAFMVSSAAGVMAG 1405
Qy 39 ----- 38
Db 1406 AGGNYAANAFLDALAEERRAEGLPALAVANGLWEDASGLTAQLTDDTRDIRRGGLRA 1465
Qy 39 ----- 38
Db 1466 ISAEHGMRLFDNASRHSPEVPLVAAPMEPVDAEVPALLRSLHRPNVYRAALAGGAQWLAA 1525
Qy 39 ----- 38
Db 1526 LAPEERAKALLKVVRTOATAVTLGHADARTIPVTGAFRDLGIDSLTAVELNGLAKVTGLR 1585
Qy 39 ----- 38
Db 1586 LPATLVFDYPTPAVLAARLGEFTGENPVLVRTASVVGODEPLAIVGMACRLPGGVSSPE 1645
Qy 39 ----- 38
Db 1646 DLWRLVESGTDAISGFPADRGWDAESLFPDPAVGKSYCVGEGFLDASAESFDAGFFGIS 1705

Qy 39 ----- 38
Db 1706 PREALAMDPOORLIMEVSMEAFERAGIEPGSVRGSDTGVFMGAYAGGYGAGADLGGFAAT 1765
Qy 39 ----- 38
Db 1766 ASATSVLSGRVSYFFGLGEPAITVDTACSSSLVALYQAGYALROGECSLALVGGVTVMAT 1825
Qy 39 ----- 38
Db 1826 POSFVEFSRKSGLDSDGRCKAFADSADGTGWAESGVGVLVVERLSDAQAKGHQVLAVVRSS 1885
Qy 39 ----- 38
Db 1886 AVNQDGASNGLSAPNGSPSQOVIRQALANAGLTAAEVDVVEAHGTGTLGLDPIEAQAVIA 1945
Qy 39 ----- 38
Db 1946 TYQDRERPLLGLSKSNIGHAQAAGSVGIKVMALQHNTVPTLHVDEPSRHVDAA 2005
Qy 39 ----- 38
Db 2006 GAVELVRENQWPCTDRPRRAGVSSFCVSGTNAHVILESAPPAQAEAEQVETPVVASD 2065
Qy 39 ----- 38
Db 2066 VLPLVISAKTPALTEHEDRLRAYLAASPGVDTRAVASTLAIVTRSVFEHRAVLGGDVT 2125
Qy 39 ----- 38
Db 2126 GTAVSDPRVVVPFGQGMWLGMSALRDSIVFAERMAECAPALRPFVMDLFTVLDDP 2185
Qy 39 ----- 38
Db 2186 AVDRVDVQPSWRMVSLLAAVMQAAGVRPDAVIGHISQGEIAACVAGAVSMRDAARIV 2245
Qy 39 ----- 38
Db 2246 TLRSEIARGLAGRAMASVALPAQDVVELVDGAMIAAHNGPASTVIAGTPEAVDHVLTAA 2305
Qy 39 -----ELLRD----- 43
Db 2306 EARGVRRITVDYASHTPHVELIRDELIDITSSSSQAPVVPWLSTVDGWSVDSPLDVE 2365
Qy 44 ----- 43
Db 2366 YWYRNLRPEVGFHPAVGLOAEGDTVFVEVSASPVLLQAMDDVVVTVATLRRDDGATRM 2425
Qy 44 -----P-SLG--- 47
Db 2426 LTALAAQAVVHGVTVDWPAILGTATTTRVLDLPTAFQHORYWLRSDVRAAADGHPLLGTW 2485
Qy 48 ----- 47
Db 2486 DLPASDGVVLTGRVSLATHTWLAHVRGVSLLPGTAPVDLVVRAADEVSDVDELVIE 2545
Qy 48 ----- 47
Db 2546 TPLLLPQTGGVQLSVSVGGADESGHRAVMVFSQADNTDTWTRHVTATVSTSDSTVSLPEF 2605
Qy 48 -----AQ 49
Db 2606 ASWPPAQPAPVSVADFYDRLAAAGTEYGFAPGLOAQRWDRDGTVYAEVYLAEEOQEAAR 2665
Qy 50 FRVH----- 53
Db 2666 FAVHPALLDAAMHASVLTHTPDQOQSVRMPFSWNHVQIRATDTAMLRAVAATPTDQWSVR 2725
Qy 54 ----- 53
Db 2726 VADDTGRPVATIGSLVTRPVTADTLGSAADDLLTLVWTEIPTPQOSSLSVGRVEDLADG 2785

Db	4946	AICDVSDRAGLARI	LAGVSPHEPLTAVIHTAGALDDGVVSESLTAQOLETVLRPKADGAWH	5005
Qy	91	-----	-----	90
Db	5006	LHELTRDADLA	AFVMYSSAAGVLGSGGQNYAAANAFDALAEORRGEGLPALAVAWGLW	5065
Qy	91	-----	-----	90
Db	5066	EDASGLTAEMTDT	DRDRIRRGGLRAISAGHGMGLLDAASRHGEPVLLAAAMPVREAEVP	5125
Qy	91	-----	-----	90
Db	5126	ALLRLLHRPARRAA	STGSSVQWLARLAPVEREKALLKLVCDGAATVLGHADASTIPAT	5185
Qy	91	-----	-----	90
Db	5186	AAFKDLGIDSLT	AVELRNSLTAKTLRLPATLVFDYPTTALAARLGEWFGETVPVVRT	5245
Qy	91	-----	-----	90
Db	5246	SVSVVAQDEPLA	IVGMACRLPGVSSPEDLWRLLESGLTDAVSGFPTDRGMDVENLFGPAA	5305
Qy	91	-----	-----	90
Db	5306	GDSYRLQGGFLD	AAAGFDASFFGISPREALAMPQORLVLEVSWEAFAFERAGIEPGSVRGT	5365
Qy	91	-----	-----	90
Db	5366	DTGVFMGAYPGY	GIGADLGGFATASAVSVLSGRVSFFGLEGPATVDTACSSSLVAL	5425
Qy	91	-----	-----	90
Db	5426	HQAGYALROGECS	LALVGGVTYMATPQTVEFARQGLAGDRSKAFADSADGAGFSEGV	5485
Qy	91	-----	-----	90
Db	5486	GULLVERLSDAQ	HQVHLMLRSLAVNQDGASNGLTAPNGPSQORVIOAALSNAAGLAH	5545
Qy	91	-----	-----	90
Db	5546	EVDVVEAHGTG	TTLGDPIEAQALLATYGDREOPLLLGSKVSNLGHQTQAAAGVGVIMV	5605
Qy	91	-----	-----	90
Db	5606	MALQRFVPTLHV	DEPSRHDWSAGAVALTENQNPMDGRARRAGVSSFGISGNTAHV	5665
Qy	91	-----	-----	90
Db	5666	ILESAPPTQADNA	VIERAPEWLPMVISARTQSALTEHEGRLRAYLAASPGVDMRAVAST	5725
Qy	91	-----	-----	90
Db	5726	LAITRSVFEHRA	VLLGDDTVTGAATDPRVVFVFGQSORAGMGEEELAAAFVVFARIHQ	5785
Qy	91	-----	-----	90
Db	5786	QVWDLDDVPDL	EVNETGYAQPALFALQVALFGLLESKWVRPDVAVCHSVGELAAAGVSGL	5845
Qy	91	-----	-----	90
Db	5846	WSLEDACTLVSAR	ARLWQALPAGGVMMVAVPVSEDEARAVLGEVETAAVNGPSSVVLSDG	5905
Qy	91	-----	-----	90
Db	5906	EAAVLOAEGLGK	WRKTRATSHAFHSARMEPMLEEFRAVAEGLTYRTPQVMSAAGDOLTTT	5965
Qy	91	-----	-----	90
Db	5966	EYVVRQVROT	VRFGQVASYEDAVFVELGNDRSLARLVDGVAMLHGDHEAQAASALAH	6025
Qy	91	-----	-----	90
Db	6026	YVNGVTVDMPAL	LGDAPATRVLDLPTYAFQHQRYWLEGTDRMAAGGHPLLGEAVAVPGSD	6085
Qy	91	-----	-----	90
Db	6086	GVLITGRVSLATH	PWLADHAVRGSVLLPGTGFVELVVRAADEVGCDVIDELVIETPLLLP	6145
Qy	91	-----	-----	90
Db	6146	QTGGVQLSVAVA	ESDDSGRRRAVTVFSRADNVDTWTRHATVTVNSDSTLTLSPDLASAMPPA	6205
Qy	91	-----	-----	90
Db	6206	QAQPVNVADFYD	QLTVVAYEYGFQGLQAARWGDGVHAEVALAEBOAQDAARNVHPA	6265
Qy	91	-----	-----	90
Db	6266	LLDAALHAGILNT	PDTEQGVRLPFSMNQOVHVGSAMLRVAVTQAADGMSVRVADDTGR	6325
Qy	91	-----	-----	90
Db	6326	PVATIGSLVTRP	VTADTLGSTDDLLALTWTEIPAPOETSLTIORPEDLADADVPEVV	6385
Qy	91	-----	-----	90
Db	6386	VFTARPDSSENPL	AQTRTLTAQVLQKIQAWLAERFTDSTLIVRIGTCLAAAVSGLMRS	6445
Qy	91	-----	-----	93
Db	6446	QOSEHPGRFVL	VESDDDTLAPDQLAATVELDEPRLRISDGRFEVPRLTRTHAABPESEKV	6505
Qy	94	-----	-----	93
Db	6506	WDPDGTVLITG	SGVLAGIAARHLVAERGVNRHLLLSRSPDGALISELSELGAQOVATAV	6565
Qy	94	-----	-----	93
Db	6566	CDVSDRPGELVR	LADVSPHEPLTAVIHTAGVVDGVVESLTAQRLDVTVLRPKADGAWHLH	6625
Qy	94	-----	-----	93
Db	6626	ELTRDADLAA	FVMYSSAAGVFGSAGQNYAVANAFDALAEORRAEGLPALALAWLWEG	6685
Qy	94	-----	-----	95
Db	6686	TSGLTANLTDH	DIRIRSGMRAISAEHGMRLFDGASRRRDPVLVAAMPEVREAEVPAM	6745
Qy	96	-----	-----	100
Db	6746	LRSLHRPVARRA	ASGGAVRCLAAALAPBEERAKALVKVWCDSATVLGHADVDSIPVTAAPR	6805
Qy	101	-----	-----	100
Db	6806	DLGVDSLTA	VELRNSLTAKTLGLRLPATLVFDYPTPGALAAARLEELFTGENPVQVTRPVSA	6865
Qy	101	-----	-----	100
Db	6866	VQDEPLAIVGM	MACRLPGVSSPEDLWRLVESGTDASISGFTDRGMDVENLFDSDPDAAAG	6925
Qy	101	-----	-----	100
Db	6926	KSYCVEGGFLA	TAAAFDASFFGISPREALAMPQORLVLEVSWEAFAFERAGIEPGSVRSGD	6985
Qy	101	-----	-----	100
Db	6986	TGVFMGAFPGY	GIGADLEGYATAGLVNLSGRLSYFFLEGPAVTVDTACSSSLVALHO	7045
Qy	101	-----	-----	101
Db	7046	AGYALROGECS	LALIGGVTVMATPHTFVFSRQRGLASDGRCKAFADSADGTGHWSEGVV	7105
Qy	102	LYITRF-----	-----	107
Db	7106	LLVERLSDAQ	HQVHLAVVRSSAVNQDGASNGLSAPNGPSQORVIRQALANAGLTTAEV	7165

Qy	108	-----	107
Db	7166	DVVEAHGTGTTGLDPIEAQVIATYQDRDQPVLLGSKSVNGHTQAAAGVSGVIKMYMA	7225
Qy	108	-----	107
Db	7226	LOHGLVPRTLHVDEPSRHWDTGAVELVTENQSWPEAGRPRAGVSFGVSGTNAHVIL	7285
Qy	108	-----	107
Db	7286	ESAPPTQAVDDVRPADAPVVASVMASELVLVISAKTQSALAEYEGRLRAYLAASPGVDM	7345
Qy	108	-----	107
Db	7346	RAVASTLAMTRSVFEHRAVIGDVTVSGTAATDPRVVFVPOGSGORAGMGAEALAAFPV	7405
Qy	108	-----	110
Db	7406	FARIHOQVWDLDDVPLEVNETGYAOPALFALQVALFGLLESNGVPRDAVIGHVSGELAA	7465
Qy	111	-----	115
Db	7466	AYVSGLSLEDACLTYSARARLWQALPAGGVMAVPVSEDEARAVLGEVGEIAAVNGPSS	7525
Qy	116	-----	115
Db	7526	VVLSGDEAAVLQAAEGLKWTRLATSHAFHSARMEPMBEEFRAVAOGLTYHAPGVVMAAG	7585
Qy	116	-----	119
Db	7586	DRVMTAEYWRQVDRTVRFGEQVASYEDAVFVELGADRSRLARLDGVAMLHGDHETQAAI	7645
Qy	120	-----	122
Db	7646	GALAHVLVNGVTVDWTALLGDVPVTRVLDLPTVAFQOQRYWAEVGRSADVSGAGLDAVGH	7705
Qy	123	-----	122
Db	7706	PLIGAVVALPGSGVLLTGRVSLATHAWLADHAVRGSVLLPGTGfVELVTRAADVGCDDV	7765
Qy	123	-----	122
Db	7766	VDELIVEAPLLLPTQGGVQVSVSVEADESGHRVVTIFSRADNADTWVRHVSAVISTSDS	7825
Qy	123	-----	122
Db	7826	TVSLSEFASMPQAQPVDTGFDPLMAEGYEYGPAPQGLQAARWYGDTVFAEVALAE	7885
Qy	123	-----	122
Db	7886	QVREAAARYPSVHPALLDAAHACTLECADAEVGVGLPFSDGVRVHAGGSAMLRAVATQA	7945
Qy	123	-----	122
Db	7946	ADGWSRVADEIGRPVANVGSVLTVPYADALGSAADDLLALTWAEIPAEGTGLTVGRF	8005
Qy	123	-----	122
Db	8006	EELVSDGDPVPEVAVFTALPNDNDDPLEQTRKLTQVLOAVQEWLGGERSFDSTLVVRT	8065
Qy	123	-----	122
Db	8066	GTGLAAAGVSGLMRSAQSEHPGRFVLVESDDDLALDQALAAVGLDEPRLISDGRFEAP	8125
Qy	123	-----	122
Db	8126	RLTRTHAAPESEKYNDPDGTVLITGGSGVLGIAVRHLVTERGVRVHLLLSRSAPDEAL	8185
Qy	123	-----	125
Db	8186	INQLGELGARVETAAACVSDRAALAQVLAGVSPHEPLTAVIHTAGVLDDGVVSLTAORL	8245
Qy	126	-----	125

Db	8246	DAVLRPKADGAWNHELTRDADLAAAFVMYSSAAGVLGSGCGNYAAANAFDALABORHA	8305
Qy	126	-----	125
Db	8306	EGLPALAVAMGLMEDASGLTAQLTDTDRDIRRGGLRAISAHEGMGLFDSASRHSSEPLV	8365
Qy	126	-----	125
Db	8366	AAPMEPVDAEVPALLRSLHRPVARRAAAAGGARWLAALAPABEREKALLKLVSDGAATVL	8425
Qy	126	-----	125
Db	8426	GHADTSTIPATTAFAKDLGIDSLTAVELRNSLAKATELRLPATLVDFYPTPTTALAARDEL	8485
Qy	126	-----	125
Db	8486	FTGENPAPVRGPVSAVAQDEPLAIVGMACRLPGGVSSPEDLWRLVESGTDASIGFPTDRG	8545
Qy	126	-----	125
Db	8546	WDVENLYDPDPDAPCKSVSVGGFLDAAAGFDASFFGISPREALAMPQORLMLEVSWEA	8605
Qy	126	-----	125
Db	8606	FERAGIEPGSVRGSDTVGFIGAYPGYGIGADLGGFGTTAGAAASVLSGRVSYFFGLEGA	8665
Qy	126	-----	131
Db	8666	FTVDTACSSSLVALHQAGYALROGECSLALVGGVTVMPTPQTFFVFSRQGLSADGRCKA	8725
Qy	132	-----	131
Db	8726	FADAADGTWAEVGVLLVERLSDAQANGHQILAVVRSSAVNQDASNGLSAPGPSQOR	8785
Qy	132	-----	131
Db	8786	VIRAALSNAGLAPHEVDVVEAHGTGTTGLDPIEAQVIATYQGGGEPELLGLSKSNVGH	8845
Qy	132	-----	133
Db	8846	TQAAAGVSGVIXKMWALQHSMPRTLHVDEPSRHWVDMWSAGAVELVAENQPMETGRPRA	8905
Qy	134	-----	133
Db	8906	GVSSFGISGTNAHVILESAPOSGDGTAGTPELVSELVPLVISAKTOPALTEHEDRLRA	8965
Qy	134	-----	133
Db	8966	YLAASPGVDIRAVASTLAVTRSVFEHRAVLLGDETVTGTAVSDPRIVFVFGQGHQWLGM	9025
Qy	134	-----	133
Db	9026	GSALRDSSVVFPAERMAECAALSEFVDWDLFAVLDDPAVVDVVDVVPASWAMVMSLAAV	9085
Qy	134	-----	135
Db	9086	WQAAGVRPDVAVIGHSGEIAAACVAGAVSMDAARIIVTLRSQAIARGLAGRGAMASVALP	9145
Qy	136	-----	135
Db	9146	AQDVLDGAWIAAHNGPASTVIAGTPEAVDHLVTALRORGAGAADHVDYASHTPHVELI	9205
Qy	136	-----	139
Db	9206	RDELLDITSSSSQDPLVPLWSTVDTGTVWDSPLDGEVWYRNLRKREPVGPHPAVSQLOAQD	9265
Qy	140	-----	139
Db	9266	TVFEVSASPVLMQAMDDDDVTVTLRRDDGDATRLMTALAAQYVHVHVTVDWRAVLGDVP	9325
Qy	140	-----	139

```
Db 9326 ATRVLDLPTAFOHQRYWAEGRSADVSAGLDVAUGHPLLGAVLAMPGSDGVMLTGRVSL 9385
Qy 140 -----TGF----- 142
Db 9386 ATHAWLADHAVRGSVLLPCTGFFVELVRAADEVACDWDDELIVEAPLLLPQTGGVQLSVS 9445
Qy 143 ----- 142
Db 9446 VGEADESGHRAVTVFSRADSDAWVRHVSATVSVDTPVTSDLTAMPPAQAKVDVAGF 9505
Qy 143 ----- 142
Db 9506 YDQLTRAGYEYGPATGQLOAARWGDGTVEAVALAEETQDAARFAVHPAVLDAAHAGI 9565
Qy 143 ----- 142
Db 9566 LNTPADRDTRLVLPFSNNHVVHVTGSATLRVATRVADGWGRVADDIGRPVATIGSLV 9625
Qy 143 ----- 142
Db 9626 TRPVAADALGSAVDLFTLTWTEIPVSQQVGVTVGKFEDLADGEVMPDVMVVFTALPDSG 9685
Qy 143 ----- 142
Db 9686 DPLAQTRRLTAEVLOAVQVWLAGERTDSTLVVRTGTGLAAAAVSGLMRSAQSEHPGRPV 9745
Qy 143 ----- 142
Db 9746 LVESDDTLTPQLAATVGLDEPLRLVIDGRVEAPRLTRTGVAEPEPEGVDPDGTVLIT 9805
Qy 143 ----- 142
Db 9806 GGSVLAGIAARHLVAERGVHRHLLLSRSPADEALISELAELGAAVVDVAVCDVSDRAGL 9865
Qy 143 ----- 142
Db 9866 ARVLAVSPDHPLTAVIHTAGVLDGQVVESTARRDLTVLRPKAGAMNHLTRDIDLA 9925
Qy 143 ----- 142
Db 9926 AFVMYSSAAGVLGSAGOGNYAVANAFVDALAEQRRAEGLPALALAWGLWEDASGLTAKLT 9985
Qy 143 ----- 142
Db 9986 GTDHRIRRSGLRTITAERGMLFDIASQGEPLVATPMPREVPEVALLRLLRPVA 10045
Qy 143 -----DLGV 146
Db 10046 RRAASTGSSAQLVGLAPEERAKALLKVVRDASAATVLGHADARSIPATGAFKDLGV 10102

RESULT 12
O85168
ID O85168 PRELIMINARY; PRT: 9376 AA.
AC O85168;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Syringomycin synthetase.
GN SYRE.
OS Pseudomonas syringae (pv. syringae).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=321;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99047670; PubMed=9830033;
RA Guenzi E., Galli G., Ggurina I., Gross D.C., Grandi G.;
RT "Characterization of the syringomycin synthetase gene cluster. A link
RT between prokaryotic and eukaryotic peptide synthetases.";
RL J. Biol. Chem. 273:32857-32863(1998).
DR EMBL; AF047828; AAC80285.1; -.
DR HSSP; P14687; 1AMU.
```

```
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00501; AMP-binding; 8.
DR Pfam; PF00668; Condensation; 9.
DR Pfam; PF00550; pp-binding; 9.
DR Pfam; PF00975; Thioesterase; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS50075; ACP DOMAIN; 9.
DR PROSITE; PS00455; ACP BINDING; 8.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_5.
KW phosphopantetheine.
SQ SEQUENCE 9376 AA; 1029826 MW; F770C08975EF9CES CRC64;

Query Match 30.7%; Score 237; DB 2; Length 9376;
Best Local Similarity 1.2%; Pred. No. 27;
Matches 94; Conservative 24; Mismatches 28; Indels 7809; Gaps 29;

Qy 1 AAGG----- 4
Db 980 AAGGSEVVVPANLIPEHCDRITPQMLPLINLSQAQIDHVVVKMPGGVAVQDIYPLAPL 1039
Qy 5 ----- 4
Db 1040 QAGILYHHISAEQDPYTLKALFALSADRAQLDDPSGALQGVINRHDILRTAVLWEDLDEP 1099
Qy 5 ----- 4
Db 1100 VQVVLQAEHLHVTELLDPDGPVDEQLHQRFRDRHYRLDVRTAPLMRIVFSDHPVNDRW 1159
Qy 5 ----- 4
Db 1160 LAMLLCHHLVSDATSLSVLQEIQAHLGQGNALGEAVPYRNYVAQRLGVSEAQHEAFP 1219
Qy 5 -----ILHL-- 8
Db 1220 RDMLGDIDEPTLPFGLQDVQDRGRNLEEASVTLAEELNLRRAQARQAGSVASLHMLAW 1279
Qy 9 ----- 8
Db 1280 ARVLGNVSACEQVVGTVLLGRMQADGADRALGMFINTPLRLVDIGATTVAEGLKATHE 1339
Qy 9 ----- 8
Db 1340 RLTAALLGHEAPLVLAQRCSGVAAPTPLFSALLNRYHSVSDVRQKLPGLGSSLRHRRP 1399
Qy 9 ----- 8
Db 1400 YQLPVDNLVDDLGDAFTLTIAQVOOISATRIGEYMQVALRNLYDALEHTPQALNSLSIL 1459
Qy 9 ----- 8
Db 1460 PDDERELLLTGFNDTAHPYPRDVLHQLIEOAAAQRPDACAVRGDSGTLLTYAELNQOAN 1519
Qy 9 -----ELLVA-- 13
Db 1520 QLAHRLIELGVEPDRVAVSLRRGAEMVVALLGILKAGGAYVDPDLPDARQAYMLSDS 1579
Qy 14 ----- 13
Db 1580 APRAVLTSHELLADLPDLGVPAVLVDGRDPSALLKKQPTGNPDAKALDLPNHLAYLYLT 1639
Qy 14 ----- 13
Db 1640 SGSTGTPKGVNNEHLGVNRLHARDAYQVNSQDRVLQKTPFGFDVSVWEFFLPILTCAE 1699
Qy 14 ----- 13
Db 1700 LVMAPSGHQDPDYLAQVDSAGITLLHFVPSMLDVFLEHRSTRDFPOLRRVLCSEALP 1759
Qy 14 ----- 13
```


Db 1760 RALQRRFEQHLKGVELHNLGYPTAAIDVTAWECRPTDGDSPICRPIANIQMHVLDAL 1819
Qy 14 ----- 13
Db 1820 GOLQPMGVAGELHIGGIVARGYLNQOLSAERFIADPFSNDQARLYKTDGVRWLANG 1879
Qy 14 ----- 13
Db 1880 ALEYLGRNDFQVKIRGLRIEIGEIEAALAKHPAVHEAVWTAREDPGDKRLVAYYTQAE 1939
Qy 14 ----- 13
Db 1940 HTAVDLETLRHLQOVLPEYMPAIYVLEAMPLTNGKLDKALPAPDGDALISRGVEA 1999
Qy 14 ----- 13
Db 2000 POGEEIQAVIWDLLGVEQVGRHRDHFELGGHSLAVSLIDRLRKQDLNLTNTVFTA 2059
Qy 14 ----- 13
Db 2060 PSVREMAIAISOEKOALFRAPANRIPAHCTQLTPDMPLVLVELSVAQIELIASAVPGGAAN 2119
Qy 14 ----- 13
Db 2120 IQHIYPLAQEGILFHYLLNERDAYLVRSTIEFDSRARLDAFLEGLQTVIDRHVLR 2179
Qy 14 ----- 13
Db 2180 SVHWVGLPOAVQVVRHQASPIHTLTPRDEDAISOLDRLSDPGRRLDLRQAPLLLAYI 2239
Qy 14 ----- 13
Db 2240 ARDPDSERWLLALIDHHMISDHVTLLEIEIRLLMRGOSAEALLPPQYREFVAQTASP 2299
Qy 14 ----- 13
Db 2300 SSAHEAFTGRADVDSPTAPFELLEVOGDNDVEESELASSDLCAIRIQARQMS 2359
Qy 14 ----- 13
Db 2360 AVLPHVAWAQVLARCTGRDDVFGTAVTGRLOGLGAERAMGFMFMTLPVRVQLATQSVQ 2419
Qy 14 ----- 13
Db 2420 ELVMATHRDLSELLSHEQASIALAQRCSVATGVPLFSSLLNYRHQEDSOLQWPGRL 2479
Qy 14 ----- 13
Db 2480 DSAERTNYPLCLSVNDYGSIDLGLIHSVQPADPORLCAAMQCALEQLTDALHTPQKEVT 2539
Qy 14 ----- 13
Db 2540 QLDVLPAAERNLLLETFTNQTRQDYPDTLCIHLFEAQVTRTPDAIAVAFOAQRLSYAE 2599
Qy 14 -----VGP----- 17
Db 2600 ROANRLAHLILGLIGDPRVAICVERGVEMVGLLGLVKGAGAYVPLDPAYPAERLAYM 2659
Qy 18 ----- 17
Db 2660 IEDSQAALMTQRHLOEYPLTLPLVLLDDQDKTFTERDDNPVVEALGVRNLAYVIYT 2719
Qy 18 ----- 17
Db 2720 SGSTGNPKGVMIEHRGLVNYSDAARLFDLSPTDVTVLQNTLNFDLSVEEIPFALLAGAT 2779
Qy 18 ----- 17
Db 2780 LTPSREIFGSEGTEHNGINPTVLHLLTAHWHTLVAEWHKQPVAEQRLQHVRLINVTGDA 2839
Qy 18 ----- 17
Db 2840 LSAOKLKLWDEVRPAHTRLLNTYGPTEATVSTAAVYSHDAAAGSESGNATIGKPMANT 2899

Qy 18 ---VFOAHQE-----D 25
Db 2900 RIVLLDAHQQPVPYGVAGEIYIGDGVARGYLNLEEVNAERFLADPFSSEPDARMYKTD 2959
Qy 26 TERYV----- 30
Db 2960 LARYMADGRIEYLRNDFQVKVGRFRIELGEIARLGNCTGVKEAVVIAREDNPGDKRLV 3019
Qy 31 -----LT----- 32
Db 3020 AYVVAQPSQLTAADLRAELAPQALAEYMLPSAFVLLDELPLTPNRKLDKRALPAPDADAL 3079
Qy 33 -----NLN 35
Db 3080 ISRGYEAPOGETETQIVAIWQDLLGIEQVGRHRDHFELGGHSLAVSLIERLURKOGCLN 3139
Qy 36 I----- 36
Db 3140 VKTVFTAPTREMALAIISODKHVLFQVAPNRIHAHTQLTPDMLPLVELSVAQIELITSA 3199
Qy 37 ----- 36
Db 3200 VAGGAANIQDIYPLAPLODGIILFHYLLNRERDAYLMRSMIEFDSRARLDAFLEGLQTV 3259
Qy 37 ----- 36
Db 3260 RHDILRSSVHMIGLPOAVQVVRHQALPVHTLTLPTEEDALSOLDRLSDPGRRLDLRQA 3319
Qy 37 ----- 36
Db 3320 PLLLAYIARDPNSERWLLALIDHHMISDHVTVVELIEIRLLMRGOSADLLPPQYRDFV 3379
Qy 37 ----- 36
Db 3380 AQTASPSSAHEAYFTRRLADVDSPTAPFELLEVOGDNDVEEAKLNSDLCAIRITQA 3439
Qy 37 -----GAE----- 39
Db 3440 RERGHSPAVLFHVAWAQVMARCTGRDDVFGTVTGRLOGTAGARAMGFMFMTLPVRVQ 3499
Qy 40 ----- 39
Db 3500 LTTQGAQELVMATHRDLSELLSHEQASIALAQRCSVATGVPLFSSLLNYRHQEDNRLO 3559
Qy 40 ----- 39
Db 3560 WPGMRLDGTERTNYPLCLSVNDYGSSELDLIHSMQPANPORLCAAMQCALEQLTDALAH 3619
Qy 40 ----- 39
Db 3620 TQMAVTOLDVLPAAERNLLLETFTNQTRQDYPDTLCIHLFEAQVTRTPDAIAVAVOGOR 3679
Qy 40 ----- 39
Db 3680 LSYADLNQANRLAHLILSLGIVPDDRVAICVERGVEMVGLLGLVKGAGAYVPLDPAYP 3739
Qy 40 -----LSLGAQ- 49
Db 3740 AERLAYMITDSQPAALLTLPLGLQDRLPALSMPLVLLDDEQVGLAECDNDNPVPTLGVN 3799
Qy 50 ----- 49
Db 3800 LAYVIYTSGSTGNPKGVMIEHRGLVNYSDAARLFDLSQSDTVLQNTLNFDLSVEEIPF 3859
Qy 50 ----- 49
Db 3860 ALLAGATLAPREIFGSEGTEHGIQTVLHLLTAHWHTLVAEWHNQPVAAEQRLQHVRL 3919
Qy 50 ----- 49
Db 3920 INVTGDALSQOKLKLWDEVRPAHTRLLNTYGPTEATVSTAAVYSDAAAGSESGNATI 3979

QY 50 ----- 49
Db 3980 GKPMANTRIYLLDAHQQPPYGVAGEIYIGGDCVARGYNLEEVNAERFLADPFSESDA 4039
QY 50 ----- VKMVL-
| | | | |
Db 4040 RMYKTGLARYWADGRIEYLGRRNDFQVKVGFRIELGETEARLGNCKGVKEAVIAREDN 4099
QY 61 ----- 60
Db 4100 PGEKRFVAVVAQPOQTITAAELRAELAPOLAEYMLPSAFVLLDELPTTPNKKFDRKALP 4159
QY 61 ----- TEPEGANIT-ANLTSSLSV-
| | | | |
Db 4160 APADDAFASREHVEFGATEIALAQIWQSLDLERVRHQDFELGCHSLLAMRLISQAR 4219
QY 81 ----- 80
Db 4220 HQGVVELGLAALFAHPEVSALAVATAQAGRSKFPDIPVARDQAWPLSFGQORLWFLAQM 4279
QY 81 ----- 80
Db 4280 EGASAAHYHIPAGLSLHGNLNLKALORALERIVARHEGLRTTFMQGDDGQPVORISPADTG 4339
QY 81 ----- 80
Db 4340 FNLQMHDLQGLADAEKLOALASEESLSQFDLQOQPLIRGRRI RMAEDHHVLLTLHHIV 4399
QY 81 -CWS-
| | | | |
Db 4400 SDGWSVDVLTRELSALYAFSODDDPLAPLEQLDYAVWQRRMLSGDVLOQQSNFWQO 4459
QY 85 ----- 84
Db 4460 TLADAPALLMLPTDRARPALQYAGAAALPVVFXDLTRGLKALSORRGSTLFWTVMAAWA 4519
QY 85 ----- 84
Db 4520 GLLGLRAGQDDVVIGTPVANRRTRSEVEGLVGLFVNTLAIKRVDSLDPKPTAETLLARVKQT 4579
QY 85 ----- OTINP-
| | | | |
Db 4580 LDAQAHQDLPFQEVQVEVINPRLSLSHSPIFOAMLSWENNEASDLTLGDMTLKSLIADT 4639
QY 90 ----- 89
Db 4640 AQFDLTDMAEVDQLVGTLEYATALFDESTMRRYLYGFQVLEAMVADDDQVLEHVALL 4699
QY 90 ----- 89
Db 4700 GADEREHLLAGLNATEAPFPQDRTIHQLFEERVQAQBDATAVAFGAQRLSYAELNRQANR 4759
QY 90 ----- EDDT 93
| | | | |
Db 4760 VAHHLISLGIKPDRDVAICVERGVEMLIGVLGVKAGAAVPLDPAYPAERLAYMIEDST 4819
QY 94 ----- 93
Db 4820 PSALLAQRDVQAHPLTLDLPVLLDEQRTTLSEDDNPVVTLAGHNLAYIYITSGTG 4879
QY 94 ----- 93
Db 4880 VPKGVVHEHRLGFAVSAANEQLYALHAPLNHLQWAGFFDVFSADLIRSLAFGGTLVLCPR 4939
QY 94 ----- DP-----GHADLV-
| | | | |
Db 4940 ETLMDPPALYRLLEEISIGFADFVLAVLNALLGWEEETHGDLRFMRVTVCSDIWAHSA 4999
QY 103 ----- 102
Db 5000 ROLRKLCDHVQVQVQAYGVTEASIDSTCFEATSQVDGVLPIGRALANTRIYLLDELQ 5059
QY 103 ----- 102

Db 5060 AVPEGVAGELYIGGAGIARGYLNLPOLTAERFMDNPFVAGERLYRSCDMARYRADGNIEF 5119
QY 103 ----- YITR-
| | | | |
Db 5120 LGRNDSQAKLRLGLLELGEIEARLAEVAGVRESLVVIREDSGGTPKLIAYFEVATRDES 5179
QY 107 ----- 106
Db 5180 GPALTPRALRQQLNLPEYMIPAAFVRMAALPLSANGKLDRRALPEPDADAFDOHDEA 5239
QY 107 ----- 106
Db 5240 ADGPLETAIAIADVLGVAQVGRHDDFFALGGHSLLMVRVLAQVROOLNLEVSFSVFA 5299
QY 107 ----- 106
Db 5300 APVURQFAERLGNTOQDARNAIKPVQRSGALPLSYAQORLWFLAQLEGSSAAHYH PAGLR 5359
QY 107 ----- 106
Db 5360 LRGNLQASLQALDRIVARHEALRTTFVQEQGPAPQRISAETGFRLOQLQVLAGOTDA 5419
QY 107 ----- 106
Db 5420 EDKLLATAAQEASEHFDLVNGPLVRGRLVNRANDNHVLLVTMHHILSDGWSGTGVNLREL 5479
QY 107 ----- 106
Db 5480 ALYAPFROGAGDPLPALPVQYVDYALWQDMLSGDVLOOQRYWQOQALAGAPALLTLPTD 5539
QY 107 ----- 106
Db 5540 RPRPAQDYSQOTLELVDTLTQTRGLKALSQRHGSSLFMTVMGAWAALLGRLSQDDVVI 5599
QY 107 ----- 106
Db 5600 GTPVANRMSAEVENLIGFFVNTLAIKRVDSLSTPSVEALVRQVKORTLAAQANQDLPFEQV 5659
QY 107 ----- FDLE-
| | | | |
Db 5660 VEVQOPQRSLSHSPITFOAMLSQNNEDTALVGLDITLQGVAVAGDTAKFDLALDIDGEVDG 5719
QY 111 ----- 110
Db 5720 QLIGTLEYATALFDESTMRRYRGYFLRLLEAMAADDQVLEQVPLLDTAEREYLLKDINA 5779
QY 111 ----- 110
Db 5780 TERTYPVQQLMHLRFEAHAEAAPOAIAVRQGEOTLTYAELDSRANALAQHLPKHGVGPGT 5839
QY 111 ----- 110
Db 5840 RVAIULDRSVELLASMLATLKCGAAYLALDRLAPEERLRFMLEDESEAIMLLSRSDLTAPD 5899
QY 111 ----- 110
Db 5900 MTPRLDLDLLELSALNQBPVVPATEVAGETPACIIYTSGSTGVPKGVIVTHNGIIVRLVQD 5959
QY 111 ----- 110
Db 5960 NGYDYSABDRVAFSSNPAFDASTPEIWCALLNGCSVIEEPQVLLPEVPVAPFALLKRGCV 6019
QY 111 ----- 110
Db 6020 TAMISSTALFNLYAGLIPALAGLRWIMCGGERADPASFRRVREHSAQVRLFNGYGPTEG 6079
QY 111 ----- 110
Db 6080 TTCATSYEIFDVLPTDLSLPIGKNANVRVYVLDARREPVPVGVGDIYIGGTGVALGYL 6139
QY 111 ----- LPDGNRQ-
| | | | |
| | | | |

Db	6140	NRPELTARFSEDPFSQAGARLYRTGDLARWLPGDNLEYLARNDGQVKVGRFRVELGEI	6199
Qy	121	-----	120
Db	6200	ESVLHLYDGVNRVVVAHEASPGDTRLVAYYTVHAGVAAPOPEDLRAQLSADLAEMYWPS	6259
Qy	121	-----	120
Db	6260	VFVRLDALPLTLNGKVDQKALPVPDLNAMFERTYVAPEGATEQALAEIFQELLGLERVR	6319
Qy	121	-----	120
Db	6320	HDGFFELGCHSLLAAOLVSRVROQLNGDMALRQLFNHPTVAELAKVVDGLQTVDSDSIEP	6379
Qy	121	-----	120
Db	6380	IERNAPALSFQORLWFLDRLDPGASSAYHMPMSLLLRGELDHPRALKAALDRLVARHES	6439
Qy	121	-----	120
Db	6440	LRTTFELHGEQPVQVIAADSGFALAEDDLRSQPYEQASLNASRIADSEAAAPPDLRQGP	6499
Qy	121	-----VTO-----	123
Db	6500	LIRGRLLRADDEHMLLIIOHHIISDCWSGVLINEFTALYQAFTEQRPDPLPALSIOYA	6559
Qy	124	-----LGGACST-----	131
Db	6560	DYAAQWRRTTGERLAEQADLWREHGGGA--PTLLSLPTDRPRPVQSYRGAVPVTIDA	6617
Qy	132	-----NSCLIT-----	137
Db	6618	ALHQLERFCQAHNVTLFMGLLSAWSVLMTRLGNERDVVIGVPSANRCRTETENLIGFFV	6677
Qy	138	-----	137
Db	6678	NALALRVDLTQNPSVAQLLEQVRQTTLAAHEHQDIPFEQVTEALQPPRSMGSHPLCQVAL	6737
Qy	138	-----	137
Db	6738	SLNWTSGGELKLPGLSLHPVLQNHETAQFDMLTLASENGALNCVIEYASDLFDRSTVE	6797
Qy	138	-----	137
Db	6798	RFAQHFHTLEAMVEDVAQPVGLGPLLSPAORLASPALLQPKAVFASGLMVHQRFQFAA	6857
Qy	138	-----	137
Db	6858	AHPONIALVFGRHEVSQALNRRNRLAHELLAQGVRPDDRVAIIAERGTOICAVLAVL	6917
Qy	138	-----	137
Db	6918	KSGAAVPLDPTYPTERLGYLLTDSAPVALLAQSAQCLDVLPAHSVPLVTLGADNDSADLA	6977
Qy	138	-----	137
Db	6978	DESVDNRNPEPALCLSREHLAYVIYTSGSTGLPKGVLVEHGNVARLFDATAGQNFQGHED	7037
Qy	138	-----	137
Db	7038	WVTPFHSFAPDFSWEIHWGALCYGGKLVIVPSEVARSPDDFYALVCEQQVTLNQTPSAF	7097
Qy	138	-----	137
Db	7098	RQFTQARERSQBEHAUREVVFGGEALDRSLQPWARTPLSRTRLVNNYMGITEITVHATY	7157
Qy	138	-----	137
Db	7158	YPIQSSEIDTAMPSLIGPALDDLCRLIDDYQQPVPGVNGEIIYIGGAGVARHVLNRTEL	7217
Qy	138	-----	137
Db	7218	NAERFIADPVALOSGARLYRTGDVAHYRSDGVVNVGRNDSQIKIRGFRIELGBIEAQLL	7277

QY	138	-----	137
Db	7278	ACPEVREAMVILREDRPGDKRLVAYLIAEDGAAPESALLRSQLASVLAEHMLPSAFVTL	7337
QY	138	-----	137
Db	7338	TWPLTTNGKLDRAALPAPQOPATVSRDYEAPLGAIETTLLAAAWQELLGVERVGRDHFPE	7397
QY	138	-----	137
Db	7398	LGGHSLVISLIERLRIGLLLDVSTFVSAPTLQAAVAVLGAGTAAERVAPANLIIPVDCT	7457
QY	138	-----	137
Db	7458	ALTPDMLPVKLTROLEHIVADVPGGVANVQDIYPLSSLOEGLPHHLLQSEGDAYLMR	7517
QY	138	-----	137
Db	7518	TLATFDSRALLDKFLGALQVINRHRDIMRSSLRWQLPOPQVVHROAOLPVIQLDTAPG	7577
QY	138	-----	137
Db	7578	EDALQMLRERTNYHMRDLQOAPLIAAYITYDTROEKWLMALLDHHLLSDNVTLRIMG	7637
QY	138	-----	137
Db	7638	EIQAVMDGRADALPPSQPYENFIARACVSAEHEAYFRQLLGDVDTTTPAYGVLDVRGG	7697
QY	138	-----	137
Db	7698	DATILRSVDLSDDLARSIHSTARAQGVPTSVLFAAAGLVAAATSGRDDGIEFTVLSGR	7757
QY	138	-----	137
Db	7758	SOGTSGANHALGMFINTLPMIRLQONSVRDIVODAYQOLSGLLTHERAPLAAQRCSAV	7817
QY	138	-----EDTG-----	141
Db	7818	DASLPMFTVLNCRHGDVNTAGENIEDMGEGVHFLGSETRTNYPYIEIAVANEANGF	7877
QY	142	-----	141
Db	7878	SLTAOSIDGIDPHRTIAAYLGOAVELVNALEODPARLASLEVIPOAERQLLLDNDFNTA	7937
QY	142	-----	141
Db	7938	SDFAPAVPIHALPEDQVRNPDVALVYEDRQLSYQLNRRANHVARQLQLGVQPDERV	7997
QY	142	-----	141
Db	7998	AICAERSLDMIAGLLGVKSGAAYVPIPAHPADRMFMLQDSQPRALLTQGALSUPVGD	8057
QY	142	-----	141
Db	8058	TPMLLLDSAESLLAADDQAFDANPVVDGLTAENLAYVITYSGTSGOSKGVWVEHRSVFNF	8117
QY	142	-----	141
Db	8118	WNVLTTRTHQCPTPATVALNAGFFDMSIKGISQLFSGHKLVIIPQLLRANGSELLDFL	8177
QY	142	-----	141
Db	8178	EAHQVHAFDSTPSQDTLLSAGLLERSYQPVSVLLGGEAINASTWEKLRNCPTIRLYNM	8237
QY	142	-----	141
Db	8238	YGPTECTVDATIDLIRDLGKPSIGRPIANQVHVLDARCEPAPLVAGEIHIGGSQVAR	8297
QY	142	-----	141
Db	8298	GYLNRDELSAERFIPVDFPSDAANARLYKTGDLGRWLADGTLEYMGNRNDFQVKVRGRIEL	8357

```
QY 142 ----- 141
Db 8358 GEIENVLLAVPGIREVVIARNDSDQSRLVAVCGESVAAEHLRSELLRHLPYMW 8417
QY 142 ----- 141
Db 8418 PSFVOLDALPLTANGKLDRLRALPAGQDALASKVVEAPQGETEBAIEIWKALLHLDQV 8477
QY 142 ----- 141
Db 8478 GRNDGFEELGHHLLAVOLLRLRLKLGTRITLRELFDAPTVRGLASLVNAAAPGEAQSI 8537
QY 142 ----- 141
Db 8538 PRANRSGRLPLSPSQORLWFLDHLHDAAGAAVHLPNALRLTGALDTAALEATLDRLVARH 8597
QY 142 ----- 141
Db 8598 ETLRTRELFVDPGEVPQKIAPDSRLPLKQDLRLDLSGERTSTLARLQENATQLFDLTK 8657
QY 142 ----- 141
Db 8658 GPMRLGHLRLVADAHVLLITLHHIVSDGWSNSVLAEQVSALYAFSQGKQDPLPALPQ 8717
QY 142 ----- 141
Db 8718 YVDYAAWQSQSLDGPALQAIQDFWRKHLECAPSVLNLPLDRPRPAIQSYTGGVVEHVFPSP 8777
QY 142 ----- 141
Db 8778 ALSADLRAPSAQGSTPFVLLAGWSMLSHLSEQTQDVVVGTPVANRQHPLEPMIGFEA 8837
QY 142 ----- 141
Db 8838 NTLALRVATDRETRLNMLDRIKSLTLAAYNHODLPFEQVVSALQPTRNVSHSPLEFQVML 8897
QY 142 -----FDLGVTI 148
Db 8898 SLDNTPPSLQLPDLVELLDSEHHTTQFDLSLSL 8932

RESULT 13
Q9L4W3 PRELIMINARY; PRT; 11096 AA.
ID Q9L4W3;
AC Q9L4W3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE NYS.
GN NYS.
OS Streptomyces noursei.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1971;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11455;
RX MEDLINE=20334850; PubMed=10873841;
RA Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA Valla S., Zotchev S.B.;
RT "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT deduction of the biosynthetic pathway.";
RL Chem. Biol. 7:395-403(2000).
DR EMBL; AF263912; AAF71776.1; -.
DR HSSP; P25715; IMLA.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR004410; FabD.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acl1_transf; 6.
DR Pfam; PF00107; adh_zinc; 1.
```

```
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt C; 6.
DR Pfam; PF00550; dp-binding; 6.
DR TIGRFAMs; TIGR00128; fabD; 6.
DR PROSITE; PS00075; ACP_DOMAIN; 6.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW phosphopantetheine; Transferase.
KW SEQUENCE 11096 AA; 1150415 MW; 776CAEAFCAE551DD CRC64;

Query Match 30.5%; Score 236; DB 2; Length 11096;
Best Local Similarity 1.1%; Pred. No. 51;
Matches 99; Conservative 18; Mismatches 31; Indels 8882; Gaps 30;

QY 1 AAG----- 3
Db 384 AAGVIKTMALQGVLPRLSLHATEFTTDDVMTAGSVDLLDETVAWPGTGRARRAGVSFG 443
QY 4 ----- 3
Db 444 ISGTNAHVILEQAPTAPEEPTTEPTVRPAVVPWALSARTAAALDAQRLATGHLADTPDA 503
QY 4 ----- 3
Db 504 DPLDVGVALADGRATFEHRAVLLPDGTELAHTAGEGCAVLFSGQSGSQRCMGRHLAR 563
QY 4 ----- 3
Db 564 FPFVFAAFDEITALLDTHLDRPLREVVGWTDADLLNDGTGWAQPALFAVEVALYRLVASLG 623
QY 4 -----GILHLE----- 9
Db 624 VTPDFVGHSIGELAAAHVAGVLSLEDACTLVAARARLMQALPRGAMLAIRATEDEVTP 683
QY 10 ----- 9
Db 684 HLTDDVSIAAVNGPTSVVVGTEEAIAIGARFTAQDRKTRRLRVSHAFHSPMDPLMAE 743
QY 10 ----- 9
Db 744 FRAVAAGLTYHEPRIPVLSNLTGTVAADVADLCSADYVVRHVREAVRFADGTALTDRGVT 803
QY 10 LLVAVGPD----- 17
Db 804 TLVELGPDGVLSAMAQESLPDGAIAVPLLRKDRPEELSAVTGLARAHVGVTVRWAGLFD 863
QY 18 ----- 17
Db 864 GTGARRADLPYVFOHQRFWPTAARAADVTAAAGLGAADHPLLGATVELADGAGYLFTSR 923
QY 18 ----- 17
Db 924 LSVRTHPMLADHGVQGRALLPCTAFVELAVRAGDEAGCDRVEELTLAAPLVLPERGQVOL 983
QY 18 ----- 17
Db 984 QVRVGAPDAAGRRITLGI FSRVEDGDFLPMSSHATGVLTAGAGAPDPTFDATVWPPSGARP 1043
QY 18 -----VFO-----AHQEDTERY----- 29
Db 1044 VDLTGAYERLAALGFOYGFAPGGLRAAWRRDTEVYAEVALPGADTDPAAFGHLPALLDA 1103
QY 30 ----- 29
Db 1104 AQHAAAYADLGAISRGGLPFAMEGVSLAAAGATTVRARIAPAGEDTIVIAVDAAGGTVL 1163
QY 30 ----- 29
Db 1164 SVDSLVSREVPAAPGACGTVHRDSL FHVEMTPLOGRPGPAPATVAVLGPDPDALADTUR 1223
QY 30 -----VLTNL----- 34
Db 1224 ATGIRTTAPRDIAALADAREGPVLDLVVTLTTTPGAPVPDAAHAHTAAVLAALAAQWLADD 1283
```

QY 35 ----- 34
Db 1284 RFADARLVLRGATDGTDPAAAAAGGLIRTARTENPGRFALLDLAPDGTGRDPETLATA 1343
QY 35 ----- 42
Db 1344 LAASHDEPDIAVRGTDVHAARLARVPLATEPTTWNPDGTVLITGGTGGGLAVLARHLVAT 1403
QY 43 ----- 45
Db 1404 HGVHLLASRRGPAADGADDLTAELTGLGATVHIAACDVADPAALADLLGTVPAGHPLT 1463
QY 46 ----- 45
Db 1464 VVHTAGVDDGVGLSLTPQRLDVLRLPKADAAWHLHEATHRLDLDALFVLFSSVAATLGS 1523
QY 46 ----- 45
Db 1524 PQQANYAAGNAFLDALAARRAATGLPATSLANGPWTOSVGTSSLSLDLVERIARSGMPP 1583
QY 46 ----- 45
Db 1584 LTLEOGTALFDAALAAGPAALAPVRLDLPVLRTOGDIAPLLRGLIRTPVVRTAAQVSQTA 1643
QY 46 ----- 50
Db 1644 DGLAQLAGLDAAARREALLELVRTOIAQVLGHADATEVETGRQFQDLGFDLSLTAVELRN 1703
QY 51 ----- 50
Db 1704 ALNTATGLRLPATWVFDYPTPHALADHLRDELLGTEAESTAVPVRTRTAGTDPPVIVG 1763
QY 51 ----- 53
Db 1764 MACRYPGIASPEDLWRLVSGADATGFFPTNRGMWLDNLYDPPDRPGRTHVRAGGFLH 1823
QY 54 ----- 53
Db 1824 DAGSFDADFFGMSPREAMATDSOORLLELSWEAVERAGIDPASLRDSGTGVFAGVWYND 1883
QY 54 ----- 53
Db 1884 YGTTLTGDEYEAFRNGSAPSVASGRVSYTLGLEGPATVDTACSSSLVALHWAQAALRA 1943
QY 54 ----- 53
Db 1944 GECSLALAGGVTVMSTPSTFVEFSRQRLAPDGRSKAPAEAADGVAMSEGVGMLVLERQS 2003
QY 54 ----- 53
Db 2004 DAVRNGHEILAVRGSVNOGASNGLTPNGPSQQRVIRQALASGGLSTADVDVAEAGH 2063
QY 54 ----- 58
Db 2064 TGTTLGDPIDIAQALLATYGRDRDENPPLLGSIKSNIGHTQAAAGVAGVINKWAMRHGV 2123
QY 59 ----- 62
Db 2124 LPQTLHVDAPSSHVDWSVGAVELLTEQTAWPETGRARRAGVSSFGISGTNAHVIEQSPT 2183
QY 63 ----- 62
Db 2184 AVPATPASADRSVEEPPAVPWALSGKTPDALRDOAARLLAHVEAHPALRPVDISYLIAT 2243
QY 63 ----- 62
Db 2244 RTAFDHRVAVLGTDRAEALRALTAAGETDPAALTGTVTRGTAFLEFSGOGSORLGNR 2303
QY 63 ----- 62
Db 2304 VLYERFPAPAEALDVTALDAELGHPLURDIWGEDAQLVDRGTGTQPALFAIEVALFRL 2363

QY 63 ----- 62
Db 2364 LEAWGITPDFVAGHSIGEIAAAHVAGVLSLGDACRLVVARAVLMOSLPEGGMIAVQATE 2423
QY 63 ----- 62
Db 2424 DEVLPLLTDDVSIAAVNSPTSVVSGYENATLAVARHFDQGRRTTRLRVSHPHSPMLA 2483
QY 63 ----- 62
Db 2484 PMLDDFRAVVESLITFTAPTTPVVSNLTCELAPAEALCSADYVVRHVREAVRFPADGIRTLA 2543
QY 63 ----- 69
Db 2544 DRGVTTFVELGPDPSVLSMAQESAPEGAGTIPLLRRDRPEQAVLAALCHLOLVGEADM 2603
QY 70 ----- 69
Db 2604 SATFRGLDPVRVDLPTAFQHRWFPAARPARPDDVRAAGLGAABHPLLLGAAVQLPDDDG 2663
QY 70 ----- 69
Db 2664 ALFTGRLSLRTHPWLADHTVLGTVLLPGTALVELAVRAGDETSCHLEELTLAAPTUPE 2723
QY 70 ----- 69
Db 2724 DGATLLQVRVGSADDTGRTVTVHARPDDTADRTWTLHATGVLATTPPAAAAFDTTVWPP 2783
QY 70 ----- 69
Db 2784 ADAEPLTTDDCVAHFTTHRFAYGPAFOGLRAARAGDVLVYAEVALPESATDEAAAFGLHP 2843
QY 70 ----- 69
Db 2844 ALLDAGLHAALLADDDTGLTGLPFSWEGVTLHAGATALRVRLAPNPNGLSVTAADPAGNP 2903
QY 70 ----- 72
Db 2904 VATVTRLLARPDLDAEQTLTHSALTTRDALPHLDWTPVPLPDTANSAPPALLGPDATVLADA 2963
QY 73 ----- 80
Db 2964 LGDPAVARHATLDDLLAGDTTPPTVPLGAPLDGDTAQAHALHTRLSALTILVQOHWLATD 3023
QY 81 ----- 80
Db 3024 RLADSRVFEVTHGAVATDDAPPTDLAAAAMVGLIRSAQENPGTFTLLDLOTEPDSTTAL 3083
QY 81 ----- 80
Db 3084 SRALTLDPEQLLLRAGRARAARLTPTPTTTHTPWSADGTVLVTGTTGGLGLVARHL 3143
QY 81 ----- 82
Db 3144 VRSCGVHLLTTSRSGVGAAGAAGLVAELES LGARVVAACDVGSGSAVAELVAGVSESY 3203
QY 83 ----- 82
Db 3204 PLASVVAAGVLLDDGVVGSLTPTERLAALVRPKVDGAMNHEATRGDLDDAFVVFSSVAGV 3263
QY 83 ----- 87
Db 3264 FGAGQANYAAGNAFLDALMVHRVAGGLPGVSLANGAMDOGVMTAGLTERDVRRAAESG 3323
QY 88 ----- 87
Db 3324 MPLLTVDOGVALFDAALATGSAALVPVRLDLAALRTRGDIAPLLRGLVRAPLRRRTAATGL 3383
QY 88 ----- 87
Db 3384 ATGADTGLVORGLRDLHAORHEALLDMVRSSAALVLGHADGNDAIDAEARFRLDGLFDSLTA 3443
QY 88 ----- 87

Qy 125 ----- 124
Db 6744 PDGEPTRTGVPFSSWRGVFLAAGSASSVRVRLGRDSDGTLAIADTTGAPVASVOALSM 6803
Qy 125 ----- 124
Db 6804 RTVSVTLSATAGLARDALFRLDWSAPEPACQPDVTVTIPAVAVGTETSETSELTA 6863
Qy 125 ----- 124
Db 6864 ALRAAGADVVRTTLSTDEPAPALIALPLVASDQTGTAEAAAPVPAVHDLTRRALALVQT 6923
Qy 125 ----- 124
Db 6924 RLQEOHFADTKFVFVTRGATVGRDVAAAAMGLVRSQAQENPGCFALVDLDPDGAVGAAA 6983
Qy 125 ----- 124
Db 6984 LVAALVSGEPQLAVRGDVLRLVRRPLTEVGAGADGTGCGVGGSGVSFSGEGAVLVT 7043
Qy 125 ----- 124
Db 7044 GGTGGLGAVLARHLVAEYGVDRDLLLVSRGERAVGAGELVAELAGVARGVRVACDVDR 7103
Qy 125 ----- 124
Db 7104 AAVVELVGHAVSAVVHAAGVLDGCMVGALTGERLSAVLRPKVDAMVHLHEATRGDLDA 7163
Qy 125 ----- 124
Db 7164 FVVSSLAGVFGSGOANYAANAFDALMTRRRRAEGLPGLSLANGPWSLTDTGTCMLAD 7223
Qy 125 ----- 124
Db 7224 ABADRLTRSGVPLTAEQGLALFDAALATGATCVPRLDLSALRAQGEVPPLLRSLIRG 7283
Qy 125 ----- 124
Db 7284 RSRRAAAESATATGLRLVGLNPVERQEVLLDLVRGOVALVGLHADADDVHPARAFRE 7343
Qy 125 ----- 124
Db 7344 LGFDSLTSVELNRNLNTVTGLRLPATMVPDPTVEVLVSYLDELGLTDAEVATVQPAAV 7403
Qy 125 ----- 130
Db 7404 AVADDPIVIVGMACRYPGGVASPDLLRLVTDCGVDAVSPPTNRGWDVESLYHPDPHLG 7463
Qy 131 ----- 130
Db 7464 TSYTRSGGFLHEAGEPDPGFGMSPREALATDSQORLLLESSWEAIERAGIDPVSURGR 7523
Qy 131 ----- 130
Db 7524 TGVFAGVMYSYDYSAMLASPEFEGFQSGSSPSLASORVAYTLCLEGPVAVTDTACSSSLV 7583
Qy 131 ----- 130
Db 7584 AMHWAMQALRSCEGLALAGVTVMTSTPAVFDVDFARQGLSPDGRCKAFADAAGVWSE 7643
Qy 131 ----- 130
Db 7644 GVGVLVLRQSDAVRNGHEILAVVRGSVAVNODGASNGLTAPNGPSOORVIRQALASGGLT 7703
Qy 131 ----- 130
Db 7704 AGDVVVEAHGTGTLGDPIEAQALLATYORDREPERPLLLGSVKSNLGHTQAAAGVAGV 7763
Qy 131 ----- 130
Db 7764 IKQVLMRHWGVVPTLHVDAPSSHVDHSEGAVELLSQAAWAPETGRVRRAGVSSFGISGT 7823
Qy 131 ----- 130

Qy 112 ----- 111
Db 5664 WDLNLYDPDPAHAGTSYARTGGFLHDAADFADDFGMSPREAMATDSQORLLLESSWEA 5723
Qy 112 ----- 111
Db 5724 IERAGIDPLTRDSRTGCVFAGVMYSGYGTRLDGAEEFGOGSALSASGRVSYTFGFE 5783
Qy 112 ----- PDGN 115
Db 5784 GPAMTVDTACSSSLVALHLLAAQALRGCECTLALAGGVTVMSIPDTFIEFSRQRLAPDGR 5843
Qy 116 ----- 115
Db 5844 SKPFESADGVGMSGVGMLLERQSDAVRNGHQILAVVRGSVAVNODGASNGLTAPNGPS 5903
Qy 116 ----- 115
Db 5904 QORVIRQALASGGLSTADVADEAHGTGTLGDPIEAQALLATYGRDRDPENPLLLGSIK 5963
Qy 116 ----- 115
Db 5964 SNLGHQAAAAGVAGVIKWMMRHWGLVPRSLNITEPSSHVDWSAGAVELLTEQTAMPETG 6023
Qy 116 ----- 115
Db 6024 RARRAGISSFCISGTHAVILEQPEAAARHSAPEEADTAEAAAKAPATAHLVMPWALSGK 6083
Qy 116 ----- 115
Db 6084 TPEALRAQARLLAHLQORPELAPADIALSLATORSOFTHRVAVLSTDRDEATRLSALA 6143
Qy 116 ----- 115
Db 6144 TTAASDPALTGTVTMGRCAVLFSQGSORGLMGRELYERFPVFAEALDVVIDHDAALP 6203
Qy 116 ----- 115
Db 6204 AOAGLREVMGDDVELLNETGTOPALFAIEVALFRLVSWGVRPDFVAGHSIGEIAAAH 6263
Qy 116 ----- 115
Db 6264 VGVFSLBEDACRLVAARATLMOALPAGGAMTAVQATEDEVIPHLTDEVIAAVNGPISVV 6323
Qy 116 ----- 115
Db 6324 ISGAEATQVAQHFADQGRRTTALRVSHAFHSPMLMLAEFRAVAEGLSYATPTLPVVS 6383
Qy 116 ----- 124
Db 6384 LTQVATADELCSAEYVVRHVREAVRFADGVTALEAEGVTFLELGPDPGVLAAMARETVA 6443
Qy 125 ----- 124
Db 6444 DDTVTVPVLRNMPERTLLTALGRHTTGTPIDWAALLAPTGAPVDPDLPYAFQHRFPW 6503
Qy 125 ----- 124
Db 6504 PSGPRDTADAAVGIAGASHPLLNGIVELADEGLLFTGRLLSLQSHPLADHAVMGQVLL 6563
Qy 125 ----- 124
Db 6564 PGTALELALRAGVEGDHVEELTAAAPLVLPERGAVQTVRVGVADTTGRTVTIHSR 6623
Qy 125 ----- 124
Db 6624 PARATTTSDTHTGTDTPWQHTAGVLVAGLPATATVPFDATVPWPAHAEPVDLADFVAS 6683
Qy 125 ----- 124
Db 6684 RAGEGFGYGAFOGLRAAMRRDGEVFDVADVALPEAGRTEAEAYGLHPALLDAGLHAALWA 6743

```
Db 7824 NAHVILERPEAARRPVMTNETNVPSTVPWVLSGKTPALRAQAALLSSIEERPELRVD 7883
Qy 131 ----- 130
Db 7884 VGMSLVTRSTFHRVAVLAADRAARALSAIADEADAATAATGRVGRHAVLPFGOG 7943
Qy 131 ----- 130
Db 7944 AQRLGMRGLYERFPFAEALDVVVDHDAALPAQAGLRVVMWGDDAELLNETGTQPAL 8003
Qy 131 ----- 130
Db 8004 FATEVALFRLVESGWVPDFVAGHSIGEIAAAHVAGVFSLEDACRLVAARATLMQALPAG 8063
Qy 131 ----- 130
Db 8064 GAMIAVOATEDEVTPHLLTDDVAIAANGPNALVVSVEDAAVEIGARFAAEGRRTRRLHV 8123
Qy 131 ----- 130
Db 8124 SHAFSPLMDPLAEFRVVAEGLSYAAPSLPVVSNLTGQVATADELCSAEYWRHVREAV 8183
Qy 131 ----- 130
Db 8184 RFADGVTALEAEGVRTFLELPGDGVLAAMAGASLTSSLAVPLLRKDRPEEPAALAAQ 8243
Qy 131 ----- 130
Db 8244 LHTAGARVDMVPLFAGVGAGRVELPTYAFORGWFWPVRGVGVDGVCAGVGLSAGHPLLG 8303
Qy 131 ----- 130
Db 8304 AVELAAGAGVLTGRLSLSSHGLADHAVMGRVFPVGTALLEVMVMRAGDEVCGRVEEL 8363
Qy 131 -----TWSC----- 134
Db 8364 TLAAPLVLPGRGVVQVAVDAFDAAGRRGVGVVSCPDGVGQAVMSQHAVGLASGVADQ 8423
Qy 135 ----- 134
Db 8424 VGGFGDGVWPPQOGAVSDAEGCYELPADAGFGVPGVFOGLRAVWRGGEELFAEVALSDE 8483
Qy 135 ----- 134
Db 8484 VAESADTATGFLHPALLDASLHASLLSLEGOSADGCPALPEAWEGVSLFASGATALRV 8543
Qy 135 ----- 134
Db 8544 RLAPAGEHAVSVTAVDPTGAPVISIDALTRRLTLDEVNASHTQLSDALFGVQWTTVPST 8603
Qy 135 ----- 134
Db 8604 PAADHPSVAIIGTDHLGLAEALSSSAGATTTTAAAYESLDALIAAGPEVSPDVTLLG 8663
Qy 135 LITEDT----- 140
Db 8664 LTTEDAIAQVYNDHDTAVAGGTIGACAAAADAARLTAELRTIQAWLADERLAARLV 8723
Qy 141 ----- 140
Db 8724 FVTRGAADGQDVAAAQVGLVRSQAOTENPGTFLGLDLDTGTEASTAVLGEALTSDEPQLL 8783
Qy 141 ----- 140
Db 8784 RDGHLHAARLRLASPADTAVPTENNADGTVLITGGTGLGAQPARHLVDRYGVNRLLLV 8843
Qy 141 ----- 140
Db 8844 SRRGPDAPGTTELVAELTAHGAEVAVQACVDGDAVAALVAGVDPDEHPLRAVHTAGVL 8903
Qy 141 ----- 140
```

```
Db 8904 DOGVIGSLTEERLATVLRPKADAAWHLHEATRGDLDFVVFSSVAGVFGGAGONYAAA 8963
Qy 141 ----- 140
Db 8964 NAFLDALMAQRAAGLPGLSLAWGPDQGTGWTGMLSDAEADRLARSIGPPLSABOGLAL 9023
Qy 141 ----- 140
Db 9024 FDAALALAGTSTPDRAAGSAAASTGTGTIAIPAAALVAVPRLDAAALAAQGEVPAILR 9083
Qy 141 ----- 140
Db 9084 GLVTRTRTRTAGSVTVAGLVNRLSGLTADERRQELLELVRTQAAVLGHADPASVDST 9143
Qy 141 ----- 140
Db 9144 AQFRDLGFDLSLTAVELNRNLSTATGLRLTATLVDPYNTDALAEHLRDELFGAVESEVRV 9203
Qy 141 -----GFDL----- 144
Db 9204 PVQALPPTADDPVVVVMACRPPGCVTSPEDLWRLVDAGTDAITFTPTNRGWDLESYDP 9263
Qy 145 ----- 144
Db 9264 DPAHLGTSYTRSGGFLHEAGEFDPAFFGMSPREALATDSQORLLLESWEAIERAGIDPL 9323
Qy 145 ----- 144
Db 9324 TLRGSATGVFAGVMYSDYGSILGGKEFGFGQGSAGSVASGRVSYTLGEGPAVTVDTA 9383
Qy 145 -----GVTI 148
Db 9384 CSSSLVALHAAQALRAGECTLALAGGTV 9413
Qy 145 -----

RESULT 14
Q18559 PRELIMINARY; PRT; 7829 AA.
AC Q18559;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Hypothetical 874.1 kDa protein.
GN C41A3.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Bentley D.;
RT "The sequence of C. elegans cosmid C41A3.";
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U41541; AAK18894.1; -.
DR HSSP; P25715; 1MLA.
DR InterPro; IPR001227; AC transferase.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
```



```
Db 2011 QDMMESSILAMLSMAAIVAGRVYIFGCGPSVTIDTACSSSLVALEMAINALLDNRC 2070
QY 17 ----- 16
Db 2071 SKVIVAGVNLILNEKQGLRTNGKMLSQHMSLSFDSRAGYGRSDCCVVLMELEAKPNF 2130
QY 17 ----- 22
Db 2131 HYMTTQSVNVNHGGRSVSLTAPNGVAHKMLTTSVINQSPSLAIDYWEAHGTGTPLGDPI 2190
QY 23 ----- 22
Db 2191 EFNLTSSILONIIIGSVKASLGHGEASAGTCGLLKLFLMLTYQVPTLIHFHVLNKDINA 2250
QY 23 ----- 22
Db 2251 GSIRLPIIGEDSELVSAGISSFGVSGTNAAAIAFNNDNNKLEPIPIHKYIILPISAKNOI 2310
QY 23 ----- 22
Db 2311 SLDNLEKQILSVIPLTDVPICNTASALANNRSHFTIRNALIVNSGIVNSKMGKPHRVA 2370
QY 23 ----- 22
Db 2371 KDRYHVHKLCDLLDASLLOYDVINETYTVASLKNQSPFAMKFAIIKELTSLEYEIVA 2430
QY 23 ----- 22
Db 2431 SDGEELLAVLLANGSLKWNFNKMTIELPIGSLTTEFADHDLNSTTSSIKSYQTQPESH 2490
QY 23 ----- 22
Db 2491 NLDSPMELMKLIMKLYITGYDVDMATVYSPVEQFIALPNYQFNKQTLWFEELEIVDHYL 2550
QY 23 ----- 22
Db 2551 IGTIDESEDTLLKNOISELRHPQFKGKPLDVGTMSEITAIEALKIRNEIPFSIQNLKT 2610
QY 23 ----- 25
Db 2611 ELITLTKPAMLETNVTNRNEDEGFNVSAIYDQORLFSLNASSVEIQNIIEVPAVEVQIPDK 2670
QY 26 ----- 25
Db 2671 VVYLKECPNAVIRRRHNMVYVDSRAEQSPRTANIVLNEIIGFAPTSDMFIILGVLP 2730
QY 26 ----- 25
Db 2731 VHYMVQVDDGALMQFQMSODKRVLSNIYVLKDAKGLEIPTIRMHKKSTLLSSQEASIVA 2790
QY 26 ----- 25
Db 2791 AKTLQMAVRHKVCLAVGDVIESGLDIDESQSTGFSGLDIDSLATVDLLNRLNQYFPEI 2850
QY 26 ----- 25
Db 2851 ELTTSDLFDNPSIIDLSIMIEQLLNEKGITEPSEPNTPKTSLGRKLSIPAVRAQVLAQI 2910
QY 26 ----- 25
Db 2911 EFVENYNSKQEVQAEAPSSSSSECSNHLBESDATVDRTEIRRKVSLAVFLATETLSAED 2970
QY 26 ----- 25
Db 2971 LQSGFTELGMSLSIVDFVNLNDKYFPDDIETASDIFDYPTVDELSDHIVRKKSSVP 3030
QY 26 ----- 25
Db 3031 PAASEIMKETWNGISTVDAEHTKLENLSOSFMLENQNSINPTLKMWSNQITKLVKPS 3090
QY 26 ----- 25

Db 3091 DGNFELFELNANGOEKEIQHFTGPNNIIIDLKGFHEGSTETLYMSLLNLVKSISKLEIQ 3150
QY 26 ----- 25
Db 3151 CRFGVQSEFGLGNSISRPFMTKVAEKNPLISFAWYQNVQVSPVDSPTSITGNWLTGG 3210
QY 26 ----- 25
Db 3211 LSGIGLEIGKFIANGAENVILISRRQPTAKALRDLSTELTHIGLARKTIVLKLKNIS 3270
QY 26 ----- 30
Db 3271 AKLIVFQSFKLSFSTKVKYISDLFSKKVTFYNNLQSKCFSKHFHWFLEVLVLIKQWTV 3330
QY 31 ----- 43
Db 3331 HTIAADINDKEKLRELTKLVNGITGIHISAGVLKDSKIERQNKESFNQVPTPKANGPHV 3390
QY 44 ----- 43
Db 3391 LEEIEKHFNKYIENFIMSSFTAACNEGQLNYGVSNAYLEYQVORRRROCKSGCAIOWG 3450
QY 44 ----- 43
Db 3451 NWIDTGMATDENVRKFLANLQFLOHNDKALKYLACILTRPELIMVANIDNVLKNRK 3510
QY 44 ----- 43
Db 3511 DLPKDLINTGILPPEDFTGKMNESEFPPLSNGDFEKVSNFSEVEDEEVELELIKEVSSIL 3570
QY 44 ----- 43
Db 3571 MCSPTKLKNNKINMDGLDLSKLIVEFLNFINSTFKISVNLSDAYNHPTLEKLAHI FQOM 3630
QY 44 ----- 43
Db 3631 TIVDHPVNSVKEEIFKSTDFCPIFINIFPNKNDPDKAKSTAVKLLENGSQLTAGKY 3690
QY 44 ----- 49
Db 3691 AVSVVKGSIIRDVVSKIKEAAPQIKLCOESSKCVMLTQGSQOYPMMGROLVENYEIPRT 3750
QY 50 ----- 49
Db 3751 TLQSLKKCDEYLOGDVSLWEILFNTDHYKLLQLTQHMOPIMFCFGYATAQLWSLGIVP 3810
QY 50 ----- 50
Db 3811 DYYLGHVSGELVAGVLGIMSIEDGLRLIVERGKAMENIAGLALLAVOREIADEVLRKF 3870
QY 51 RV-----HLVKM-- 57
Db 3871 KVS VATINSKQVVFAGTKSVLDAALAFVKGOGKOATYVNOOYPFHNSLIQETHLVSLRQ 3930
QY 58 ----- 57
Db 3931 CLADIKFSAGRTPLVSNVTGQIINTFSEAYIVKHTVSAVKFVDCVETLOAKGVTVWIDAG 3990
QY 58 ----- 57
Db 3991 SAAVLATFVKRIIQPTLSKHRIVOTCKEKESDVNLVOACLEBOSGLPISWTTLYGCG 4050
QY 58 ----- 57
Db 4051 RNADERLVEFPMTHNDIIKNDEBELLEGHQLNGKIWAGAYQFKIDQLVKLKAAGMELM 4110
QY 58 ----- 57
Db 4111 LKNVKFLKPWYIEDNREYQIOWNSDMTIELIVNSVIVCSLEVEPQNSVLKLETISENEKP 4170
QY 58 ----- 57
Db 4171 FEVHDFYETLFRNGLOYDSGFRRIESARRSKRCFSQIKSSPFAWPLIDSAMHSITASVV 4230
```

Qy	58	-----SOTINPEDDT-----	84	-----SOTINPEDDT-----	93
Db	4231	PRRDCYFLPVMGSMVTMKTNSFTLPNLHAQTVITSETDKFIOVNALLAGDTPICEVR	57	-----SOTINPEDDT-----	93
Qy	58	-----VILTEP-----	94	-----DP-----	95
Db	4291	NMTIVLKTPEVHTRIPIINSIETVETPPKSEIEIVGFDISLPYNQISENSEWHLKNTV	96	-----DP-----	95
Qy	64	-----	96	-----DP-----	95
Db	4351	KOKLHNRSLKODHARVALLDSARYDPEYFGRPSEAKFIDPOORLLCGSAKLDSLL	96	-----DP-----	95
Qy	64	-----	96	-----DP-----	95
Db	4411	ITSLTNTGVFIGCSANEFSHIVAYGYKOPRAEWSGGTNSALAGRIAHWLKLGPVVT	96	-----DP-----	95
Qy	64	-----	96	-----DP-----	95
Db	4471	LDACSSSFYALSACDALRTGQCEYALVGTVNLVHMTTDLQNAKMTVDVDFCKAFDV	96	-----DP-----	95
Qy	64	-----EGAPNITA-----	96	-----DP-----	95
Db	4531	DANGYKRSEAVCSMLLTSPNIDSVATITNYATGHNGTSSSLFTPNGLSQLEVMMQATNP	96	-----DP-----	95
Qy	72	-----NLTSSLS 79	96	-----DP-----	95
Db	4591	LEKILEIQHTCTGKLGDPPIEINAIKSLVSSACKIGSVKSNIGHTGSSGLVSLCSSLS	96	-----DP-----	95
Qy	80	-----	96	-----DP-----	95
Db	4651	FRSKYVAQLHLKCPETNSIKTNKMICRFIGEDADENNILINFGFTGSCNCSVVLKPKNA	96	-----DP-----	95
Qy	80	-----	96	-----DP-----	95
Db	4711	ISEHPVSSEVFYPILLSSHSAKSLQYQVVLCEFISSAKSLHDIMMSLFQKKIHVHRQF	96	-----DP-----	95
Qy	80	-----	96	-----DP-----	95
Db	4771	IIFNPKRIAVTSLDGFVEVRDERLEKLKPCSVFLKEGVVHFDKDKSFQORVDLPISVN	96	-----DP-----	95
Qy	80	-----VC-----	96	-----DP-----	95
Db	4831	NTLHWALSYRDEIDRHESQMSFKNIYFVKVLETPMQPNPISKVVCIGRLDLIPKIE	96	-----DP-----	95
Qy	82	-----	96	-----DP-----	95
Db	4891	IDEVSSFANGIIVFHPISINSIFEYLKAKVMSLISRNQNVFIIICCPENGTSHTEWTGL	96	-----DP-----	95
Qy	82	-----	96	-----DP-----	95
Db	4951	RLASEKMIPIYKFVSIDKVDALKLEFNHEYMFEAFYKRGYVERLKRKNPTFKFAPQY	96	-----DP-----	95
Qy	82	-----	96	-----DP-----	95
Db	5011	EKCLISGGTGIGCSAIINELKPKSSVITRKNIASDGGKTFLLSSDITRLDISHKENYVFH	96	-----DP-----	95
Qy	82	-----	96	-----DP-----	95
Db	5071	LAGIVNLSLHENVKRDSDLDMVSIKLOKAKNLMKCCDETSHFVFSSTIANVLGSYGSNY	96	-----DP-----	95
Qy	82	-----GW-----	96	-----DP-----	95
Db	5131	AFSNGLVTSFLETSSTKSTIIHGWPKVGMGLAQPERREIVKQIESNGWKLPLNODAISV	96	-----DP-----	95
Qy	84	-----	96	-----DP-----	95
Db	5191	FYQPMETHEQIIVDGDFTTIVARQPHLOKLLSEVVENTKYKEEIKKSLNFEIFFEI	96	-----DP-----	95
Qy	84	-----	96	-----DP-----	95
Db	5251	VGITDISSKLNIPFMDLGIDSLCMENLRYSLNKNFDELTVMSEMPENATYQKLOTYVETL	96	-----DP-----	95

Qy	84	-----SOTINPEDDT-----	93	-----SOTINPEDDT-----	93
Db	5311	RKAKHSESLSHDRVSSQVSNKEDDTRVAVIGWSABFSGSSNIHEYWENLMDGICSTGNK	93	-----SOTINPEDDT-----	93
Qy	94	-----DP-----	94	-----DP-----	95
Db	5371	YLLKNPFGDNKFFNLTDARVLDPOVRKFIQHAYLALENSQYKQKHELRCGVPAGAE	94	-----DP-----	95
Qy	96	-----DP-----	94	-----DP-----	95
Db	5431	PSDYGRADDDHDDMRKLFVMMNNSYLASVASYCLDLKGBAVSVSACSTALVAVANAVKS	94	-----DP-----	95
Qy	96	-----DP-----	94	-----DP-----	95
Db	5491	IQSGSMDYALVGAASIAEYSGALSGDFDOKTMTFSGVCRPFKDSGIVRGSGVCFV	94	-----DP-----	95
Qy	96	-----DP-----	94	-----DP-----	95
Db	5551	LKRYSOALLDNNDNVHFIKIDFAINNDGHSRSMFMAPNAPGOLKCMTDVLARFTNKEKERI	94	-----DP-----	95
Qy	96	-----GHA-----	94	-----DP-----	95
Db	5611	SFVECHATGTTLDDTIEMNSLRTAYSPKXKLAIGSCCKANIGHAYASGLAALVKCAKMLQ	94	-----DP-----	95
Qy	99	-----	94	-----DP-----	95
Db	5671	TGIIPQVNFSEPRDGMGOFTVNGKKTSTSONSLISDSFGIGGTNVHMFIEFPARSOE	94	-----DP-----	95
Qy	99	-----DLVLY-----	94	-----DP-----	95
Db	5731	VVKISSENILYDMIPISAKTEYSLDHTSEAISKYLOTDNTNKIAOCSFTIHSRVPMSR	94	-----DP-----	95
Qy	104	-----	94	-----DP-----	95
Db	5791	TYLSVNNNELLKIRTNKKTWNGKSPKIALFPAPOGIOFTNLPNEYLKNSAYRREVEY	94	-----DP-----	95
Qy	104	-----	94	-----DP-----	95
Db	5851	LCEASSFGIPSLGILYPTKQNFHLIHAFOAIAIFVQCMIAFKAIKNVFNPTCLIGH	94	-----DP-----	95
Qy	104	-----	94	-----DP-----	95
Db	5911	SVGEYAAVVISGALKTEALKLLIKRSELIGKTEKARMLMVMNYEKOLPSHVHVSIIIDA	94	-----DP-----	95
Qy	104	-----	94	-----DP-----	95
Db	5971	NTKCVGPVETIDNLEKYFINNHKYNRIETKHGPHSKMPCISKEFEPPCESFATKVPL	94	-----DP-----	95
Qy	104	-----	94	-----DP-----	95
Db	6031	IPMISITGSEIKIFDSKYCTMHTNPVNLVVDHIMKLDIDIIVEVGPTGVLSNLLAK	94	-----DP-----	95
Qy	104	-----	94	-----DP-----	95
Db	6091	RNSKIVVPTCGTKGHPKISLGEICIGOLMSNGVDIRKLTPLKLVADGOVPGYCFDEROFGE	94	-----DP-----	95
Qy	104	-----	94	-----DP-----	95
Db	6151	KNQISOKLNRFSYKESKWPVSELEETKPVRTVCDGSLNGISEPVLVLLRKSSEDIH	94	-----DP-----	95
Qy	104	-----TRFDL-----	94	-----DP-----	95
Db	6211	KNYFDVELFLKSPRCSTVVVGMDNSPSVHLSGLIRCYQLVSRIDLKYVENFENTAMS	94	-----DP-----	95
Qy	110	-----ELP-----	94	-----DP-----	95
Db	6271	IVIOKVLTANGLYPRIDSTGLCKHGFRTTEIPSRLHPVSRGHALVFGANGFIGSIVFRLL	94	-----DP-----	95
Qy	113	-----	94	-----DP-----	95
Db	6331	QEMGMNVIPISRASIPSCDITNICKVQNFSLGFKKFSVVINCVGVETSAMKMNKTSLEQ	94	-----DP-----	95
Qy	113	-----	94	-----DP-----	95

```
Db 6391 EIVLSPKTEGYNILKCLBEFSIEVDKLVNFSSLSVWPLLGNFDFYASANCEVETLKQG 6450
Qy 113 -----DGNQVRGVT----- 122
Db 6451 SKYIKOFLTSLPPLGSRMYESTSTXHNLSQILMFOBELTNVLKTAFAESNMNGIVFV 6510
Qy 123 ----- 122
Db 6511 SSVNPADIATKSIAPHRTDENGSEVKAVESLKLPKSTCEFEVIAEIKWETLGLISILNDAN 6570
Qy 123 -----QLQG----- 126
Db 6571 PNFFSLGGDSLSALQVWKVKQKTDRIVDVNDLFDNPTLQEFTKFVKNLTKFAAGNTND 6630
Qy 127 ----- 126
Db 6631 KISYDAIPLNTSQTQMFRLQIDTTSKYNLIFKITISYETKFWBEFLKYLHSLIAYQPS 6690
Qy 127 -----ACSPT----- 131
Db 6691 YRTVFKSGNSPYQYICSLTESFHDPKRCNLNNAISHPEHNLFEIGKSTPLRVRAEDCD 6750
Qy 132 -----WSCLITEDT----- 140
Db 6751 NSRIHIVFNQHHILTDGWSMTVLSDTVSSLYAAYRGETSFPSPKTKOTISQVAMGTSKD 6810
Qy 141 ----- 140
Db 6811 IKEALEYQNTYHTIIPYDSETGNTSPSYRISKIPSKIWQKLGLSKLYNTTMYNLAL 6870
Qy 141 ----- 140
Db 6871 SVFCDVRSFTGQADILLAYAISGRNADNSSELIGYFMNNAFKTSLPFELRLLEEILNIV 6930
Qy 141 -----G 141
Db 6931 LNSLEKSRGFATIPFYQMYEQNRKLNELSLFFNFROKLDYPTVSMFEGAKCEIEHLNLNA 6990
Qy 142 FDLGVTI 148
Db 6991 FDFSFTI 6997

RESULT 15
Q9L4X3
ID Q9L4X3 PRELIMINARY: PRT; 9477 AA.
AC Q9L4X3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE NYSI.
GN NYSI.
OS Streptomyces noursei.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1971;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 11455;
RX MEDLINE=20334850; PubMed=10873841;
RA Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA Valla S., Zotchev S.B.;
RT "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT deduction of the biosynthetic pathway."
RL Chem. Biol. 7:395-403 (2000).
DR EMBL; AF263912; AAF71766.1; -.
DR HSP; P81989; IQE5.
DR InterPro; IPR001227; AC transferase.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR004410; FAbD.
DR InterPro; IPR001899; Gram_pos_anchor.
```

```
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR001005; Myb_DNA_binding.
DR InterPro; IPR003890; Pantne_attach.
DR Pfam; PF00698; Acyl_transf; 6.
DR Pfam; PF00106; adh_short; 3.
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF00550; pp-binding; 6.
DR TIGRFAMs; TIGR00128; fabD; 6.
DR PROSITE; PS00075; ACP_DOMAIN; 6.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
DR Phosphopantetheine; Transferase.
SQ SEQUENCE 9477 AA; 988132 MW; 0D2BCA5D7B265483 CRC64;

Query Match 29.9%; Score 231; DB 2; Length 9477;
Best Local Similarity 1.1%; Pred.No. 57;
Matches 92; Conservative 27; Mismatches 29; Indels 7868; Gaps 29;

Qy 1 AAGGILH----- 7
Db 86 ASGFLHDAPDFDADFFGCSPREAVAMPQQRVVLESWEAFERAGIDPTSVKGSRTGVF 145
Qy 8 ----- 7
Db 146 IGAMAQDYRVGPADGAEGFOLTGNTGTVLSGRISYTFCTGVGPAVTVDTACSSSLVAVHLA 205
Qy 8 ----- 7
Db 206 TQALRAGECTLALAGGVTIMSGPOTFIEMGRQGLSADGRCSRSGDTADGTGWAEGVIL 265
Qy 8 ----- 7
Db 266 VLERLSDAVRNGHEILAVVVGTVAVNODGASNGLTAPNGSQOQVIOQALVNARLAAGDID 325
Qy 8 ----- 7
Db 326 VVEAHGTGTLGDPVEAQALLATYQGNRPADRPRLLLSGVSKNSLHTOAAAGVAGVIKVMV 385
Qy 8 ----- 7
Db 386 AMRHGTLPRTLHAEPTHVDSQGANVRLLTDTTDWPAATGAPRRRAVSVSGISGTNAHTI 445
Qy 8 ----- 7
Db 446 IEQAPEPEQEDAAQDDAAGSTPATAPVVPVVPVLLSGRTPDALRGQAAALRAALDTG 505
Qy 8 ----- 7
Db 506 RRPDLLDLAHSLATTPRAGFEHRAVLLATDHPALTDGLTALADADDPAAPAWITGTTTAE 565
Qy 8 ----- 7
Db 566 TRLAVLFTGQAGRLGAGRELAARPPAFATADALDAFTPHLDRLPREVLWGTDAALLD 625
Qy 8 ----- 7
Db 626 RTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAAATLVAARG 685
Qy 8 ----- 7
Db 686 RLNQALPDGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTALAEOL 745
Qy 8 ----- 7
Db 746 AADGRKTRRLRVSHAFSLMEPMLDAFRAVVEDLTLOPPLLPPVSVNLTKGPATVAQLTS 805
Qy 8 ----- 7
Db 806 ADYWDVHVHRAVRPADGIDMLARHDTTAFLELGPDGVLSSAMAQDCLDAADAVATLPAUR 865
```

```
QY 8 ----- 7
Db 866 AGRPEHTLTATLHAGLHVHGTATLDMTCFAGTGARRTDLPTAFORRYWPKALQSGTAD 925
QY 8 -----LELL 11
Db 926 LRSVGLGAAHPLLSAAVSLADAGGTLTGRSLRQHPWLADHTVRGTTLLPGTAFLELA 985
QY 12 VAVG-----PDV----- 18
Db 986 VRAGDEVGCDRVEELTLAAPLLPEQGGVQVQLWIGNPDVSGRRVTNVHARPDGTDDTPW 1045
QY 19 ----- 18
Db 1046 TAHATGVLTTADASROLPASSEQGTPLAGDPHPALDAAQWPPAGAEPLPLDGHYRLAD 1105
QY 19 ----- 18
Db 1106 GFGYGPVFOGLRAWRGGDVVYAEVELPEAGRSDAEAFGLHPALLDAAALHAAPTGLGE 1165
QY 19 ----- 18
Db 1166 RGRGLPFSMEGSLHAGGATTLRLVRLTPVADDALALTADGTGAPVLSVDSLVRSVAT 1225
QY 19 ----- 18
Db 1226 QOLDTAAAVARDALFRLDWTVPQATDPGPVALLGADPFGLLTHAGFADAPAYDPLAAL 1285
QY 19 ----- 18
Db 1286 AAADGPVTTVLSLACTGDDAADPARSAHRCAEAELAAVQTLWDHHERFAAARLVFVTR 1345
QY 19 ----- 18
Db 1346 GATVRDVAANAAMVGLVRSQSNPGCFALVDLPDGAAGAAALVAALVSGEPOLAVRGD 1405
QY 19 ----- 18
Db 1406 VLRVARLVRPLTEVGAGADGTGVDGSGVSFSGEGAVLVGTGGTGLGAVLARHLVAE 1465
QY 19 ----- 18
Db 1466 YGVRDLLLLRSRGERAVGAGELVAELAGVARGVRVACDVTDRAAVVELVGHGAVSAVH 1525
QY 19 ----- 18
Db 1526 AAGVLDDGMVGLTGERLSAVLRPKVDVAVHLHEATRGDLDAFVVFSSLAGVFGSQQA 1585
QY 19 ----- 18
Db 1586 NYAAANAFDALMTRRAEGLPGLSLANGPWEQSGGTMGTLDVDAERLARGVPPPLSVA 1645
QY 19 ----- 18
Db 1646 QGLALFDAAVAGTATCVPRLDLPVLBARGEVPPLRLSLRVRARRAAVAGSATAGNLA 1705
QY 19 ----- 18
Db 1706 QRLRLDEDEGRMDLVLRGQVALVLGHATGQVDYDAGRAPDLGFDLSLTAVELNRINT 1765
QY 19 ----- 18
Db 1766 VTGLRLPATLVFDYPTVRHLATYVLDELLGTDAEVATVQPAAVADDPIVIVGMACRYP 1825
QY 19 ----- 18
Db 1826 GCVSSPEDLWRVLTEGDAVSGFPTNRGWDVESLYHPDPDHGTSYTRSGGFLHEAGEFD 1885
QY 19 ----- 18
Db 1886 PGFGCMSPREALATDSQORLLLESWEAIERAGIDPVSLGRSRTGVFAGVMYSYAML 1945
QY 19 -----FO----- 20
```

```
Db 1946 SPEPFGQSGSGSPSLASGRVAYTLLEGPAVTVDTACSSSLVAMHMAQALRSCECGLA 2005
QY 21 ----- 20
Db 2006 LAGGVTMSTPAVFDVFARQRLSPDGRCKAFADAADGVGMSEGVGLVLERQSDAVRNG 2065
QY 21 ----- 20
Db 2066 HEILAVRGSANQDGASNGLTAPNGPSQORVIRQALASGGLTAGDQDVVVEAHGTGTTIG 2125
QY 21 ----- 20
Db 2126 DPIEAQALLATYGRORPERPILLGSKVSNLGHQTAAGVAGVVKMLAMRHGVVPRTLH 2185
QY 21 ----- 20
Db 2186 VDAPSSHVDWSEGAVELLSEQAAMPETGRVRRAGVSSFGISGTNNHVIVEQAPGAKAIAA 2245
QY 21 ----- 20
Db 2246 AGAARTPGAVPVLISGRGRSALRGQAARLLGHLQARPDABLVDVALSLATTSRFEQRA 2305
QY 21 ----- 20
Db 2306 AVVAQDRDQLIASLGLAADRDPAPVVEGEAAGRGRVAVLFTQGSORAAAMGRELHEVOP 2365
QY 21 ----- 20
Db 2366 EFAAFDAVCAVDFDPLLDRLPLREVVFAPEDGDEAALLDETGTQTOPALFAVEVALFRLVES 2425
QY 21 ----- 20
Db 2426 WGVPRDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQATEDEI 2485
QY 21 -AHQEDT----- 26
Db 2486 AAHLDDTVAIAAVNGVPSQSVVISGDEEAETIAATFAERGRKTKRLRVSHAFHSPRMDGML 2545
QY 27 ----- 26
Db 2546 DAFRIVAEGLTVRAPRIPLVSDLTGRRADDAEVCTAEYVVRHVREAVRFADCVRTLRDAG 2605
QY 27 ----- 26
Db 2606 ATTFLGSDGLLTAAEDTLGDDHDAELVPMLRAGRAEELAAATALARLOVRGVVDWA 2665
QY 27 ----- 26
Db 2666 AYLAGTGARRTDLPTVAFQHAYYWPOLPTPAALAAADPADQOLWAAVEMGDARELADIL 2725
QY 27 -----ERYVLTNL----- 34
Db 2726 GLGEQDLTPLDLSLLPALTSWRRGNQEKHLDDTLRYRVENWTRLSKPTAPVLDDGTLLVASD 2785
QY 35 ----- 34
Db 2786 ATAADQALLDGLADALSGHARGVRRLLLDDSCADRAVLAERLARTADVDAATQVLSVLP 2845
QY 35 ----- 34
Db 2846 LDERDADDCPLTRGLALTVALVOALADTGAOGLMTATRGAVSTNPADPVTHPVQAAAW 2905
QY 35 ----- 34
Db 2906 GLGRGVALEHPRLWGLVLDLPQVFDERAGORLAGILAVKADPDGEDOVALRATGVSGRRL 2965
QY 35 -----NIGAEELLR----- 42
Db 2966 VRHTEALPTAAEFTATGTVLTGGTGLCAEVARWLARAGAOHLVTSRRGPDPAGAAE 3025
QY 43 ----- 42
```

Db 3026 LRAELGCGPSVSVACDVADRDAALAAVLTAALPEELPLTGVVHTAGVGHYGPLDTLSTAE 3085
Qy 43 ----- 42
Db 3086 FAGLTAAGLAGAHLDAALADRELDFFVLFGSIAGVWGSNGOSAYCAANAYLDALALHRR 3145
Qy 43 -----DPS----- 45
Db 3146 ARGLAATSVAMGWAEAGMAADDAVSETLRRQGLGLDPAPAMTELRRVAVRQDVTVTVA 3205
Qy 46 ----- 45
Db 3206 DVDWQRYAPLFTSARPSALIAGLPEVRALAAADTEQDATGASEVVTVRVALAEPEQLRL 3265
Qy 46 ----- 45
Db 3266 LLDLVRTESATVLGHSSADAVEPGRAFRDVGFDSLTAVELRKRLGAATGLSLPSTWFDY 3325
Qy 46 -----LGAQPRV----- 52
Db 3326 PTPLELAQYLRAEILGAVLEAGVPATGADDEPIAIIQMCRFFPGGVSSPEQLMDLNAS 3385
Qy 53 ----- 52
Db 3386 GTDAISEFPVNRGMOTGHLFDPPDRPGTTYSTQGGFLHEADFEFPTFFGISPREALVMD 3445
Qy 53 ----- 52
Db 3446 POORLLLETTWESFERAGIRPETLRLSTLTGTFTVGSSYQBYGLGAGDGTGEGHMTGSSPSV 3505
Qy 53 -----HL----- 54
Db 3506 LSGRLSYVLEGPAVTDTACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNFIA 3565
Qy 55 ----- 54
Db 3566 FSRQALAKDCKAFSDDADGWTLAGVGVVLVERLSDAQRNGHPVLAVLGSAINQDG 3625
Qy 55 ----- 54
Db 3626 ASNGLTAPNGSQORVIRQALANARLAPGDIDALEAHGTGTPLGDPIDIAQALFATYGRDR 3685
Qy 55 -----VQWV----- 58
Db 3686 DPESALLLGSVKSNIQTSQAAGIASVIKMMALRHSLEPPTLHADAPSHVDWSAGTVR 3745
Qy 59 ILT----- 61
Db 3746 LLTQARAMPETGRPRRAAVSFGISGTNAHVLLLEQAPVADTPAEBRPAVAVPIAAGVVP 3805
Qy 62 ----- 61
Db 3806 WVTARSAALRGQAERLLAHAEVTGTALPAAGPLDIGLSLSARAFPHRAVAVVPPAGT 3865
Qy 62 ----- 61
Db 3866 DPLAALRAVATDGPSPVARGVADVEGRTVFVFPQGSQWGVMSQLLDES AVFAERIAE 3925
Qy 62 ----- 61
Db 3926 CAALAEFTDWSLDVLRGVGAPSLERVDVVQPASFAVMVSLAALWRSRGLVPDAVVGH 3985
Qy 62 ----- 61
Db 3986 SOGETAAAVVSGALSURDQARVVALRSQAIGRALAGRGGMMSVALSVDLPEPLVEFEGR 4045
Qy 62 -----EPE----- 64
Db 4046 VSVAAVNGPRSVVAGEPEALDALHRLTADDIRARRIAVDYASHSHQVEDLHEELLEVL 4105
Qy 65 ----- 64
Db 4106 AELAPRTSEVPFFSTVTGDWLDTARM DAGYWFNLRGRVRFADAVADLLAEYRAFVEYS 4165

Qy 65 ----- 64
Db 4166 SHPVLMSVAQBAIDEAGVPAAVAGTLRRDOGGTDRFLLSAAAEVFRGVVDVWAGLFEFTG 4225
Qy 65 ----- 64
Db 4226 ASRIDLPYAFQHEHLMWAVPPAPEAVAAAADPDDAAFTAVEDGDVSALTAAALGTDEDSVA 4285
Qy 65 ----- 64
Db 4286 AVLPAITSWRRARRDRSTVDARVRYVANKPLGGTLPHPSLTGCTWLLVTADGIDDDTDVAGA 4345
Qy 65 ----- 64
Db 4346 LETYGAEVRLVLDEECVDRVLRERLAGAEDVTGIVSVLAAAEERTDAVPGTSLVLGTAL 4405
Qy 65 ----- 64
Db 4406 TVALIQALGDAEIDAPVWALTFRGAVSTGRADELTAPOAQVGTGIGWTAALHPQWGGTL 4465
Qy 65 ----- 64
Db 4466 DLPAALDARAAQRLAAVLVSGALGSDDDQLAIRPSGVFTPRIVRABEATAGRPAGTTPRGTT 4525
Qy 65 ----- 64
Db 4526 LVTGSGGTLA PHLARWLAORGAEHLVLISRRGTAAPGAELVAELAESGTEATVAACDIT 4585
Qy 65 ----- 64
Db 4586 DRDAVALLADL KADGRTVRTVHTAATIELHTLDATLADPDRVLHAKVTGAQVLAELL 4645
Qy 65 ----- 64
Db 4646 DDEELDDFVLYSSTAGMWGGAHAAYVAGNAYLAALAEHRRANGLPALSLSWGIWADDLK 4705
Qy 65 ----- 64
Db 4706 LGRVDPQMI RRSGLFEFMDPQLALSGLQALDDNNENLAVADVDMETHYHPVVTGRPTPLF 4765
Qy 65 ----- 64
Db 4766 DEVPEVRLTAAAEOSACTVAEGEFAAALRALSDAEQOQTLLTETVTRTAAASVLGSSAED 4825
Qy 65 ----- 64
Db 4826 LTDQAFRDVGFDSLTAVGLRNLRLASVTGLTLPSTWFDYPNPAAALAYLHGLAGARSA 4885
Qy 65 ----- 64
Db 4886 AAGAAAVPTGAPDADDPTAIVGMSCRYPGVGS AEDLWRIALDEYDAISGFPADRGMDAE 4945
Qy 65 ----- 64
Db 4946 GLYDPPDRPGRYTSVQGFRLRDVAEFPDFGIGSPREALSMDPQORLLLETAMBAFEHA 5005
Qy 65 -----ITANLTS----- 75
Db 5006 GIDPVQGRGSRGTGTFVGASYQDYASGVPSNSEGSGHMITGTLSSVLSGRVSYLFGFEGPA 5065
Qy 76 ----- 75
Db 5066 VTLDTACSSSLVAMHLACQSLRNGESSLALAGGVSIMSTPMSPVGFSTRORALAEGRCKA 5125
Qy 76 ----- 75
Db 5126 YADGADGTM LAEGVGLVLLERLS DARANGHOVLAVIRGS AVNQDGASNL TAPNGPSOOR 5185
Qy 76 -----SLLSVCG----- 82
Db 5186 VIROALANS AVAPGDIDVLEGHGTGTALGDPIEAQALLATYQQDRAPERPLLLSGVKSNI 5245

QY	83	-----WSQTINP- 89
Db	5246	GHTOMASGVASVIKLVRALOEGVVPKSLHIDRPFSTHVDWSSGAIGLLTERTPWPETQRP 5305
QY	90	----- 89
Db	5306	RAAVSSFGISGTVNHTILEQAPADEAPTADPPRDGLVPVLLSGRGEAALRAQAARLLAF 5365
QY	90	----- 89
Db	5366	VEERPEAHLTDLAHSLATSRAALERRAAVIAADRDLTLRGLRALSDGRDPGLVQGTAGR 5425
QY	90	----- 89
Db	5426	GRTAFLEFTGQSORPGMGREUHQRYPVFADALDEVLARLDGDPDRPLREVLFAPDSAEA 5485
QY	90	----- 89
Db	5486	ALLDRGTGAQPALFAVEVALFRLLTSNGLTPDVLGHSVGELEAAAHVAGVLSLDDACTLV 5545
QY	90	-----EDD----- 92
Db	5546	AARGRLMQALPEGAMVALEAAEDEVLPLEGLTDRVSAAVNGPRSVVVGVEEDVLL 5605
QY	93	----- 92
Db	5606	ADLFAADGRRTKLRVSHAFSHPLMDAMLDLDDFAAVARGLTYHPPTTIPFVSNVSGGLATAE 5665
QY	93	-----TDP----- 95
Db	5666	QVTPDPYVGVHRAVRPADGIDWLATQGDVHTFLELPGDGVLSAMARESLLTDPSTALL 5725
QY	96	----- 95
Db	5726	PTLUGDRPERPALVTAIAAAHAGHARVDMGYPADHGARRTTLTPVAFQERYWPDTTAA 5785
QY	96	----- 95
Db	5786	TSAHTPGSALDAEFWAAVERDDVAALAASLDDATVTAMVPALTAWRRRRGEQTELDWS 5845
QY	96	----- 95
Db	5846	RYRVTKPRGCATAPALTCRWLVLPDHDQRODDATAAADAADVETALGTTTVRLTVIT 5905
QY	96	----- 95
Db	5906	TDRAALAARITEAAGDQGPFGVLSLLPLATGDAGHPGAPALTTLTTVAQALGDAGIDA 5965
QY	96	----- 95
Db	5966	PLMNVTGAVAVGRAEQVTAPEQAQAVNGLGRAVALELPARFGGTLDPATLDGOAARRLR 6025
QY	96	----- 95
Db	6026	AVLAATDGEDAVALRPSGVFLRLAHAPAGPDTARTAFDPAAGTVLITGTTGGIGGHVAR 6085
QY	96	----- 95
Db	6086	RLARDGATHLLTSRRGPAAPGADALRAELEELGARVTLAACDAADRDLAALLAELPDD 6145
QY	96	----- 95
Db	6146	APLCAVPHTAGVVEDHVVDALTPENFAAVLRAKTVAAHILHELTADLDLAAFLFSSTAG 6205
QY	96	----- 95
Db	6206	VLGAAGOCGYAANAHLDALEHRRSHGLTALSANGPWAGSGMWADAELTDRVRGGF 6265
QY	96	----- 95
Db	6266	EPLAPEPAVRALLRAIENDDTTVALADI DWERFORAFAAVRPLPFVADLPTETGRATPATA 6325
QY	96	-----GHAD----- 99

Db	6326	TGAATGLRQQLAELPEHERPAAVLDLLRTQVAAVLGHADPRTVEDDHAFRDLGFDLSLTL 6385
QY	100	-----LVLV----- 103
Db	6386	ELRNALNAATGLSLPATLVYDLPTPREMADFLLAELLGLTPTDTAATVASTASPKLSASF 6445
QY	104	-----ITRF----- 107
Db	6446	EOGGTFFDDPIAVIGIGCRPGVTTPEELWQLLDEGRDGISRFFDDRRGWDLAALCAGAS 6505
QY	108	----- 107
Db	6506	DTLEGGLTGVADFDARFFGISPREALAMPQORLLLETTWEALERAGIDPTTLRGSTTG 6565
QY	108	----- 107
Db	6566	VFVGTNGODYPTLLRRSASDVAGYVATGNTASVMSGRLSYALGLEGPVAVTIDTACSSLV 6625
QY	108	----- 107
Db	6626	ALHWAGRALRAGECLLVAGGVSVMASPDSFVEFSTOGGLAPDGRCKAFSDAADGTAWSE 6685
QY	108	----- 107
Db	6686	GVGILLVLERLSAARNRHQVLGLIRGTAVNQDGASNGLTAPNGLSQORVIAQALADARLR 6745
QY	108	-----DLEL----- 111
Db	6746	PADIDAIEAHGTGTLGDPPEARALITAYGRDRDAERPLLGLTVKSNIGHTQAAAGAAGV 6805
QY	112	----- 111
Db	6806	IKXLMAMRHGTLPTLHVGTPTSSHVDWGGTVALLDDARHPWPTGQPRRAGVSAPGVSGT 6865
QY	112	----- 111
Db	6866	NAHVVEQAPETEAAPAAEPAPATPTVVPVYVSGRSREALQALDRLTAHTAAHAPAR 6925
QY	112	-----PDGNRQ----- 117
Db	6926	SAADVGRSLATDRTLPHRAVLLAGPDGVREAAARAAAPRTPGRTAFLFSGOGAOLHMGH 6985
QY	118	----- 117
Db	6986	DLYQRPVYADALDTVLAQFDTVLDPVPLRAALFAAPGTPPEAALLDOTGTQPALFAVEVA 7045
QY	118	-----VRGVTLQGGACS----- 129
Db	7046	LFRLAESHWRLTPDFVAGHSIGEIAAAHVAGVFSLEDACTLVAAARASLMQPRDGMVAL 7105
QY	130	----- 129
Db	7106	EATEDEVAPLLTDGVALAANGPRSVVWAGAEADVAVADRLAADGRRTRRLTTSVSHAFHS 7165
QY	130	----- 129
Db	7166	PLMDPMLTDFARVAEGLTYHEPRIPLVSTLLGAPAGAEURTDPYVVRHVRETVPFADGVR 7225
QY	130	----- 129
Db	7226	ALHDAGAGTFVEIGPDGVLTALTQOTLDTVEAGAPAVVVPLORRDRAGDLALLEGLATLH 7285
QY	130	-----PTW----- 132
Db	7286	THGTGPSWPAYPEATGGHRTDLPTVAFQORYWYVPELGAPVATAPODPAARVYHETWAPLP 7345
QY	133	----- 132
Db	7346	APEAAAPAGRALVLPAGNRDRTAMMTAVADALGADTVTAEPDALABOLTAAGDTPWRVV 7405
QY	133	----- 132

Db 7406 SLLAAASEGLPADGAWPAALLATLDEAGVHAPLWCVTRGAVAVAGEAPTAVGQAALWGLG 7465
Qy 133 ----- 132
Db 7466 RVAALDHPDRFCGLADLPADTDAAHAAGLLAAHAAAPGTEAEIAVRAATCGHARRLVRTPAA 7525
Qy 133 ----- 132
Db 7526 ADGATWLPGTGTVLVVGGTGTGCGRAARWLVRGARHLVLTAPDGTGTTAADTEALTAE 7585
Qy 133 ----- 132
Db 7586 LAALGARITVVVDHPTAPDGFALLDGLPDDTPLTAVVYAPEADAAPGTAELSAALAPV 7645
Qy 133 ----- 134
Db 7646 TALGAALTGRPLDAFVLFGSIAGLVGRGAAEAASGAYLDAFARACDRGTPALAVAWG 7705
Qy 135 ----- 134
Db 7706 AWADLVGPSLAAHLRMNGLPVMDADTALTALSRVADGSAAEAVADVWETFAPLHHEAR 7765
Qy 135 -----LITE----- 138
|:|
Db 7766 RTALFDALPEARALAEAAARDRAKTAAGDYGRWLAEQPAADHDATLLALVTEKAATVL 7825
Qy 139 ----- 138
Db 7826 GHADHDLLEPDLPRFDLGFDSLTAVDLRNQLTAETGLTLPATLVFDHPNPAALAAHLRAQ 7885
Qy 139 ----- 138
Db 7886 LLGEASPSAAPVAAPVALGADDDAIIVGMACRYPGGVTSPEDLWQLVGDEVDVGFPT 7945
Qy 139 DTGFDL----- 144
|:|
Db 7946 DRGHDLAALAGDGPGRSATAQGGFLYDATDFDGLFCISPREALVMDPQORILLETWEA 8005
Qy 145 ----- 144
Db 8006 LERAGIDPATLRGSGTTGVFVGGSGDYRPPPEAGQWQTAQSASLLSGRLAYTFGIQGP 8065
Qy 145 -----GVTI 148
|:|
Db 8066 VSVDTACSSSLVALHLAAQALRAGECSIALAGGTV 8101

Search completed: March 20, 2003, 12:41:58
Job time : 273 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:34:52 ; Search time 37 Seconds
(without alignments)
533.002 Million cell updates/sec

Title: SEQID_1_15FUSED

Perfect score: 773

Sequence: 1 AAGILHLELLVAGPDVQ.....SPTWSCLITDGTGDLGVTI 148

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.0

Searched: 908470 seqs, 13250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 101002:

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	773	100.0	148	23	Human Von Willebra
2	773	100.0	242	23	Human aggreganase
3	773	100.0	933	23	Human ADAM-TS-like
4	773	100.0	1353	23	Human Von Willebra
5	773	100.0	1416	23	Human ADAMTS-M pol
6	773	100.0	1427	23	Human Von Willebra
7	770	99.6	203	22	Amino acid sequenc
8	770	99.6	1120	22	Amino acid sequenc
9	529	68.4	118	21	Human ORFX ORF2432
10	272	35.2	26926	22	Human titin (Conne

11	264	34.2	15281	15	AAR44929	T. niveum Cyclospo
12	238	30.8	1784	21	AAB41379	Human ORFX ORF1143
13	237	30.7	997	22	AAB73283	Human ADAMTS-7 am
14	237	30.7	1505	23	AAU72897	Human metalloproce
15	237	30.7	1686	22	AAB74944	Human ADAM type me
16	237	30.7	1686	22	AAE00913	Human 27875 ADAM-T
17	237	30.7	1686	22	AAE00934	Human 27875 ADAM-T
18	236	30.5	11096	22	AAE10129	Streptomyces nous
19	235	30.4	12199	21	AAV77180	S. venezuelae plk
20	231	29.9	1690	22	AAB86949	Human metalloproce
21	231	29.9	9477	22	AAE10144	Streptomyces nous
22	229	29.6	2150	21	AAV53898	Amino acid sequenc
23	229	29.6	2165	22	AAB90617	Human secreted pro
24	226	29.2	5532	21	AAB23752	S. avermitilis ave
25	226	29.2	5532	22	AAE65267	Streptomyces avam
26	224	29.0	6685	22	ABG21030	Novel human diagno
27	224	29.0	7107	22	ABE58144	Drosophila melanog
28	224	29.0	9222	22	ABG21064	Novel human diagno
29	223	28.8	1054	22	ABE60410	Drosophila melanog
30	216	27.9	4829	22	AAB97833	Human apoptosis-in
31	216	27.9	5588	20	AAV39301	SpnE a polyketide
32	216	27.9	5588	22	AAB70969	S. spinosa protein
33	216	27.9	7068	22	AAE10142	Streptomyces nous
34	216	27.9	7337	22	ABG22216	Novel human diagno
35	215	27.8	4260	22	AAW23830	Human EST encoded
36	215	27.8	4349	23	AAU79940	Human protocadheri
37	215	27.8	4349	23	ABB97540	Novel human protei
38	215	27.8	4854	22	ABG06027	Novel human diagno
39	214	27.7	5985	23	AAU84799	HCV HepCia acrambl
40	214	27.7	6797	22	AAB31558	Pimaricin biosynth
41	212	27.4	5035	13	AAU25450	MH mutant porcine
42	212	27.4	5546	23	AAU85008	Human melanocyte d
43	212	27.4	6619	22	ABG23329	Novel human diagno
44	211	27.3	4999	23	AAO22158	Ramoplanin biosynt
45	211	27.3	7718	22	ABG11811	Novel human diagno

ALIGNMENTS

RESULT 1

AAE24448

ID AAE24448 standard; Protein; 148 AA.

XX AC AAE24448;

XX DT 04-OCT-2002 (first entry)

XX DE Human Von Willebrand factor-cleaving protease fragment #1.

XX KW Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy: enzyme;
transgenic animal; immunisation; thromboembolic disease; preclampsia;
XX KW thrombotic thrombocytic purpura; TTP; Henoch-Schonlein purpura;
XX KW thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
XX KW transgenic; anticoagulant.

XX OS Homo sapiens.

XX PN WO200242441-A2.

XX PD 30-MAY-2002.

XX PF 20-NOV-2001; 2001WO-EPI3391.

XX PR 22-NOV-2000; 2000US-0721254.

XX PR 12-APR-2001; 2001US-0833328.

XX PA (BAXT) BAXTER AG.

XX PI Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;

XX PI Scheiflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;

XX PI Zimmermann K, Voelkel D;

DR WPI; 2002-479950/51.
XX N-PSDB; AAD39331.
PT Novel isolated or substantially purified Von Willebrand factor-cleaving
PT protease, useful for producing preparation for therapy of thrombosis
PT and thromboembolic disease such as thrombotic thrombocytopenic purpura
XX
XX
PS Claim 1; Fig 3; 93pp; English.
XX
XX The invention relates to an isolated or substantially pure Von Willebrand
XX factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
XX purifying vWF which involves providing vWF-cp as a ligand, contacting a
XX solution comprising vWF with the polypeptide, ligand under conditions
XX where vWF is bound to the ligand and recovering from the ligand purified
XX vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
XX which involves immunising an animal with vWF-cp and isolating the anti-
XX vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
XX producing a preparation of prophylaxis and therapy of thrombosis and
XX thromboembolic disease such as thrombotic thrombocytopenic purpura (TTP),
XX HemoCh-Schonlein purpura, preeclampsia, neonatal thrombocytopenia or
XX haemolytic-uraemic syndrome. vWF-cp can also be used for processing
XX plasmatric or recombinantly produced vWF. The invention is useful for
XX construction expression systems and generating transgenic animals which
XX express the polypeptide in vivo. The present sequence is human vWF-cp
XX fragment.
SQ Sequence 148 AA;
Query Match 100.0%; Score 773; DB 23; Length 148;
Best Local Similarity 100.0%; Pred. No. 3e-33;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGGILHLELLVAVGPDVFOAHQEDTERVLTNLNIGALLRDPDLGQFRVHLVKMVL 60
DB 1 AAGGILHLELLVAVGPDVFOAHQEDTERVLTNLNIGALLRDPDLGQFRVHLVKMVL 60
QY 61 TEPEGAPNITANLTSSLSVCGWSQTIINPEDDTPGHADLVLYITRFDELDPGNGRVRG 120
DB 61 TEPEGAPNITANLTSSLSVCGWSQTIINPEDDTPGHADLVLYITRFDELDPGNGRVRG 120
QY 121 VTQLGACSPWTSCLITEDTGFDLGVTI 148
DB 121 VTQLGACSPWTSCLITEDTGFDLGVTI 148
RESULT 2
AAU97641
ID AAU97641 standard; Protein; 242 AA.
XX AAU97641;
XX
XX 27-AUG-2002 (first entry)
XX Human aggrecanase protein.
XX Aggrecanase; osteoparthic; antiarthritic; antiinflammatory; human;
XX cartilage; osteoarthritis; inflammatory disease; enzyme.
XX
XX Homo sapiens.
XX WO200233093-A2.
XX
XX 25-APR-2002.
XX
XX 17-OCT-2001; 2001WO-US32458.
XX
XX 18-OCT-2000; 2000US-241469P.
XX
XX (GEMY) GENETICS INST INC.
XX
XX Racie LA, Twine NC, Agostino MJ, Wolfman NM, Morris EA;
XX WPI; 2002-452389/48.
XX

DR N-PSDB; ABK52579.
XX
XX Novel purified aggrecanase polypeptide useful for developing inhibitors
XX and antibodies to the aggrecanase polypeptide, which are useful for
XX treating aggrecanase-associated condition such as osteoarthritis
XX
XX Claim 1; Page 33-34; 41pp; English.
XX
XX This invention relates to the cDNA and protein sequences of a novel
XX human aggrecanase polypeptide. The protein of the invention may be
XX used to inhibit the proteolytic activity of aggrecanase, or to inhibit
XX the aggrecanase-mediated cleavage of aggrecan in cartilage. The
XX protein of the invention is useful for developing inhibitors of
XX aggrecanase protein. The cDNA sequence encoding the aggrecanase protein
XX of the invention is useful for designing probes for obtaining DNA
XX sequences encoding other aggrecanase molecules. The cDNA sequence is
XX also useful for detecting mRNA encoding aggrecanase in a given cell
XX population, and thus for detecting or diagnosing genetic disorders
XX involving the aggrecanase, or disorders involving cellular, organ or
XX tissue disorders in which aggrecanase is irregularly transcribed or
XX expressed. The DNA sequences may also be useful for preparing vectors
XX for gene therapy applications. An inhibitor of the protein is useful
XX in treating conditions characterised by degradation of articular
XX cartilage, by blocking the enzyme's proteolytic activity. An
XX aggrecanase protein inhibitor and a method for inhibition of its activity
XX are useful for treating various aggrecanase-associated conditions
XX including osteoarthritis and other inflammatory diseases. The
XX present sequence represents the human aggrecanase protein of the
XX invention.
SQ Sequence 242 AA;
Query Match 100.0%; Score 773; DB 23; Length 242;
Best Local Similarity 100.0%; Pred. No. 1e-32;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGGILHLELLVAVGPDVFOAHQEDTERVLTNLNIGALLRDPDLGQFRVHLVKMVL 60
DB 42 AAGGILHLELLVAVGPDVFOAHQEDTERVLTNLNIGALLRDPDLGQFRVHLVKMVL 101
QY 61 TEPEGAPNITANLTSSLSVCGWSQTIINPEDDTPGHADLVLYITRFDELDPGNGRVRG 120
DB 102 TEPEGAPNITANLTSSLSVCGWSQTIINPEDDTPGHADLVLYITRFDELDPGNGRVRG 161
QY 121 VTQLGACSPWTSCLITEDTGFDLGVTI 148
DB 162 VTQLGACSPWTSCLITEDTGFDLGVTI 189
RESULT 3
AAU79217
ID AAU79217 standard; Protein; 933 AA.
XX AAU79217;
XX
XX 15-JUL-2002 (first entry)
XX Human ADAM-TS-like protein.
XX
XX Human; ADAM-TS-like protein; cardiovascular disorder; angina;
XX vascular system; congestive heart failure; myocardial infarction;
XX ischaemic heart disease; arrhythmia; hypertensive vascular disease;
XX secondary arterial hypertension; peripheral vascular disease; embolism;
XX chronic peripheral arterial occlusive disease; acute arterial thrombosis;
XX inflammatory vascular disorder; chronic obstructive pulmonary disease;
XX liver disorder.
XX
XX Homo sapiens.
XX OS
XX WO200226999-A2.
XX
XX 04-APR-2002.
XX
XX

PF 26-SEP-2001; 2001WO-EP11124.
XX
PR 28-SEP-2000; 2000US-235881P.
PR 25-JUL-2001; 2001US-307333P.
XX
PA (FARB) BAYER AG.
XX
PI Xiao Y;
XX
DR WPI; 2002-383274/41.
DR N-PSDB; ABK493356.
XX
XX New purified human ADAM-TS-like protein, useful for identifying
PT modulators of protein activity for treating cardiovascular or liver
PT disorder or chronic obstructive pulmonary disease -
XX
XX Claim 25; Fig 2; 106pp; English.
XX
XX The invention relates to a human ADAM-TS-like protein and the
CC polynucleotide encoding it. The protein of the invention is useful for
CC treating cardiovascular disorders including diseases of the heart and
CC vascular system, such as congestive heart failure, myocardial infarction,
CC ischaemic heart diseases (e.g., stable angina, unstable angina), atrial
CC and ventricular arrhythmia, hypertensive vascular diseases (e.g., all
CC kinds of secondary arterial hypertension), and peripheral vascular
CC diseases (e.g., chronic peripheral arterial occlusive disease, acute
CC arterial thrombosis and embolism, inflammatory vascular disorders, etc),
CC liver disorders and chronic obstructive pulmonary disease. The sequences
CC are useful in diagnostic assays for detecting diseases and abnormalities
CC or susceptibility to diseases and abnormalities related to the presence
CC of mutations in the nucleic acid sequences which encode the protein. The
CC sequences are also useful for modulating ADAM-TS-like protein activity in
CC a disease condition. This sequence represents the human ADAM-TS-like
CC protein.
XX
SQ Sequence 933 AA;

Query Match 100.0%; Score 773; DB 23; Length 933;
Best Local Similarity 100.0%; Pred. No. 2.7e-31;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLRDPGLCAQFRVHLVKMVL 60
Db 78 AAGGILHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLRDPGLCAQFRVHLVKMVL 137

QY 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRG 120
Db 138 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRG 197

QY 121 VTOLGGACSPWTSCLITDGTGFDLGVTI 148
Db 198 VTOLGGACSPWTSCLITDGTGFDLGVTI 225

RESULT 4
AAE24449
ID AAE24449 standard; Protein; 1353 AA.
XX
AC AAE24449;
XX
XX 04-OCT-2002 (first entry)
XX
DE Human Von Willebrand factor-cleaving protease fragment #2.
XX
KW Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;
KW transgenic animal; immunisation; thromboembolic disease; preeclampsia;
KW thrombotic thrombocytopenic purpura; TTP; Henoch-Schonlein purpura;
KW thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
KW transgenic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200242441-A2.

XX
PD 30-MAY-2002.
XX
PF 20-NOV-2001; 2001WO-EP13391.
XX
XX 22-NOV-2000; 2000US-0721254.
PR 12-APR-2001; 2001US-0833328.
XX
PA (BAXT) BAXTER AG.
XX
XX Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
PI Scheifflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
PI Zimmermann K, Voelkel D;
XX
XX WPI; 2002-479950/51.
XX
PT Novel isolated or substantially purified Von Willebrand factor-cleaving
PT protease, useful for producing preparation for therapy of thrombosis
PT and thromboembolic disease such as thrombotic thrombocytopenic purpura -
XX
XX Claim 1; Page 64-68; 93pp; English.
XX
XX The invention relates to an isolated or substantially pure Von Willebrand
CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
CC solution comprising vWF with the polypeptide ligand under conditions
CC where vWF is bound to the ligand and recovering from the ligand purified
CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
CC which involves immunising an animal with vWF-cp and isolating the anti-
CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
CC producing a preparation of prophylaxis and therapy of thrombosis and
CC thromboembolic disease such as thrombotic thrombocytopenic purpura (TTP),
CC Henoch-Schonlein purpura, preeclampsia, neonatal thrombocytopenia or
CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
CC plasmatic or recombinantly produced vWF. The invention is useful for
CC construction expression systems and generating transgenic animals which
CC express the polypeptide in vivo. The present sequence is human vWF-cp
CC fragment.
XX
SQ Sequence 1353 AA;

Query Match 100.0%; Score 773; DB 23; Length 1353;
Best Local Similarity 100.0%; Pred. No. 6.7e-31;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLRDPGLCAQFRVHLVKMVL 60
Db 1 AAGGILHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLRDPGLCAQFRVHLVKMVL 60

QY 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRG 120
Db 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRG 120

QY 121 VTOLGGACSPWTSCLITDGTGFDLGVTI 148
Db 121 VTOLGGACSPWTSCLITDGTGFDLGVTI 148

RESULT 5
ABB04153
ID ABB04153 standard; protein; 1416 AA.
XX
AC ABB04153;
XX
XX 26-MAR-2002 (first entry)
XX
DE Human ADAMTS-M polypeptide.
XX
XX Osteoarthritis; rheumatoid arthritis; inflammatory bowel disease;
KW Crohn's disease; asthma; Alzheimer's disease; organ transplant rejection;
KW cachexia; allergy; cancer; leukaemia; lymphoma; osteoporosis;
KW atherosclerosis; congestive heart failure; myocardial infarction; stroke;
KW neurodegenerative disease; autoimmune disorder; Huntington's;

KW Parkinson's; migraine; pain; depression; multiple sclerosis; burn;
KW infertility; diabetic shock; gene therapy; ADAMTS-M;
KW A Disintegrin And Metalloprotease; thrombospondin domain.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Peptide 1..97
FT /label= Prodomain
FT /note= "The mature form of the ADAMTS-M protein is
FT processed by furin cleavage of the prodomain"
FT 94..97
FT Cleavage-site /label= Furin_cleavage_site
FT 98..1416
FT Protein /label= Mature_ADAMTS-M_protein
FT /note= "The mature form of the ADAMTS-M protein is
FT processed by furin cleavage of the prodomain"
FT 98..311
FT Domain /label= Metalloprotease_domain
FT 247..272
FT Domain /label= Zinc-binding_motif
FT 324..394
FT Domain /label= Disintegrin_domain
FT 410..473
FT Domain /label= Thrombospondin_submotif
FT 419..424
FT Domain /label= Heparin-binding_domain
FT 1099..1156
FT Domain /label= Thrombospondin_submotif
XX
PN EP1152055-A1.
XX
XX 07-NOV-2001.
XX
XX 24-APR-2001; 2001EP-0303706.
XX
XX 27-APR-2000; 2000US-200040P.
XX
XX (PFIZ) PFIZER PROD INC.
XX
XX Buckbinder L, Mitchell PG, Wachtmann TS, Walsh RT;
PI
XX
XX WPI; 2002-084275/12.
XX N-PSDB; ABA02549.
XX
XX New polynucleotide, useful in gene therapy, particularly for treating
PT or preventing e.g. arthritis, Crohn's disease, Alzheimer's disease and
PT organ transplant toxicity and rejection, comprises ADAMTS
PT polynucleotide and encoded polypeptide -
XX
XX Claim 4; Fig 2; 31pp; English.
PS
XX The present sequence represents a ADAMTS protein, designated ADAMTS-M,
XX that exhibits the characteristics of the ADAM (A Disintegrin And
XX Metalloprotease) family of metalloproteases, and contains a
XX thrombospondin domain (TS). The protein is encoded by the cDNA given in
XX ABA02549. The specification describes a newly isolated polynucleotide,
XX comprising a nucleotide sequence encoding an ADAMTS-M polypeptide as
XX given in the specification, or a metalloproteinase, disintegrin domain,
XX prodomain or its thrombospondin submotif. The polynucleotide, polypeptide
XX and agent are useful for manufacturing a medicament for treating a
XX subject in need of altering activity or expression of ADAMTS-M. The
XX polynucleotide, ADAMTS-M polypeptide and agent are useful for
XX manufacturing a medicament for treating arthritis (osteoarthritis and
XX rheumatoid arthritis), inflammatory bowel disease, Crohn's disease,
XX asthma, Alzheimer's disease, organ transplant toxicity and rejection,
XX cachexia, allergy, cancer (e.g. solid tumour cancer including colon,
XX breast, lung, prostate, brain or haematopoietic malignancies including
XX leukaemia and lymphoma), osteoporosis, atherosclerosis, aortic aneurysm,
XX congestive heart failure, myocardial infarction, stroke, head trauma,
XX spinal cord injury, neurodegenerative disease, autoimmune disorders,
XX Huntington's disease, Parkinson's disease, migraine, pain, depression,
XX multiple sclerosis, abnormal wound healing, burns, infertility or

CC diabetic shock. The polynucleotide and polypeptide are also useful for
CC diagnosing the diseases above. The polynucleotide is particularly useful
CC in gene therapy for treating the diseases cited above.
XX
XX SQ Sequence 1416 AA;
Query Match 100.0%; Score 773; DB 23; Length 1416;
Best Local Similarity 100.0%; Pred. No. 7.5e-31;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 AAGGILHLELLVAVGPDVFOAQHEDTERVLTNLTGAEALLRDPISLGAQFRVHLVKMVL 60
Db 98 AAGGILHLELLVAVGPDVFOAQHEDTERVLTNLTGAEALLRDPISLGAQFRVHLVKMVL 157
Oy 61 TEPEGAPNITANITSSLLSVCGWSQINPEDDDTPGHADLVLYITRFDLELPDGNRQVRG 120
Db 158 TEPEGAPNITANITSSLLSVCGWSQINPEDDDTPGHADLVLYITRFDLELPDGNRQVRG 217
Oy 121 VTOLGGACSPWTSCLITEDTGFDLGVTI 148
Db 218 VTOLGGACSPWTSCLITEDTGFDLGVTI 245
RESULT 6
AAE24450
ID AAE24450 standard; Protein; 1427 AA.
XX AAE24450;
XX 04-OCT-2002 (first entry)
XX Human Von Willebrand factor-cleaving protease (vWF-cp).
XX
XX Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;
XX transgenic animal; immunisation; thromboembolic disease; preeclampsia;
XX thrombotic thrombocytopenic purpura; TTP; Henoch-Schonlein purpura;
XX thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
XX transgenic; anticoagulant; chromosome 9.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Peptide 1..51
FT /label= Signal_peptide
FT 52..1427
FT Protein /note= "Mature human vWF-cp protein"
FT Cleavage-site 69..75 /note= "Furin cleavage site"
FT 224..228
FT Region /note= "Catalytical side"
FT 249
FT Region /note= "Met turn"
FT 301..377
FT Domain /note= "Disintegrin like motif"
FT 387..439
FT Domain /note= "Thrombospondin type I motif"
FT 441..553
FT Region /note= "Cysteine rich region"
FT 554..687
FT Region /note= "Spacer"
FT 688..743
FT Domain /note= "Thrombospondin type I motif"
FT 744..805
FT Domain /note= "Thrombospondin type I motif"
FT 897..952
FT Domain /note= "Thrombospondin type I motif"
FT 953..1013
FT Domain /note= "Thrombospondin type I motif"
FT 1016..1073
FT Domain /note= "Thrombospondin type I motif"
FT 1075..1131
FT Domain /note= "Thrombospondin type I motif"
XX

PN W0200242441-A2.
 XX 30-MAY-2002.
 XX 20-NOV-2001; 2001WO-EP13391.
 XX 22-NOV-2000; 2000US-0721254.
 PR 12-APR-2001; 2001US-0833328.
 XX (BAXT) BAXTER AG.
 XX Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
 PI Scheiflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
 PI Zimmermann K, Voelkel D;
 XX WPI: 2002-479950/51.
 DR N-PSDB; AAD39332.
 XX Novel isolated or substantially purified Von Willebrand factor-cleaving
 PT protease, useful for producing preparation for therapy of thrombosis
 PT and thromboembolic disease such as thrombotic thrombocytopenic purpura
 XX Claim 1: Fig 5; 93pp; English.
 XX The invention relates to an isolated or substantially pure Von Willebrand
 CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
 CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
 CC solution comprising vWF with the polypeptide ligand under conditions
 CC where vWF is bound to the ligand and recovering from the ligand purified
 CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
 CC which involves immunising an animal with vWF-cp and isolating the anti-
 CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
 CC producing a preparation of prophylaxis and therapy of thrombosis and
 CC thromboembolic disease such as thrombotic thrombocytopenic purpura (TTP),
 CC Henoch-Schönlein purpura, preeclampsia, neonatal thrombocytopenia or
 CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
 CC plasmatic or recombinantly produced vWF. The invention is useful for
 CC construction expression systems and generating transgenic animals which
 CC express the polypeptide in vivo. The present sequence is human vWF-cp
 CC protein. vWF-cp gene is located on chromosome 9.
 XX Sequence 1427 AA;
 SQ Query Match 100.0%; Score 773; DB 23; Length 1427;
 Best Local Similarity 100.0%; Pred. No. 7.7e-31;
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAGGILHLELLVAVGPDVFOAQHEDTERVYLTNLNIGALLRDPSPSLGAQFRVHLVKWVIL 60
 DB 75 AAGGILHLELLVAVGPDVFOAQHEDTERVYLTNLNIGALLRDPSPSLGAQFRVHLVKWVIL 134
 QY 61 TEPEGAPNITANLTSSLSVCGWSQINPEDDTPGHADLVLYITRFDELDPGNRQVRG 120
 DB 135 TEPEGAPNITANLTSSLSVCGWSQINPEDDTPGHADLVLYITRFDELDPGNRQVRG 194
 QY 121 VTOLGGACSPWNSCLITDGTGFDLGVTI 148
 DB 195 VTOLGGACSPWNSCLITDGTGFDLGVTI 222
 RESULT 7
 AAG63826
 ID AAG63826 standard; Protein; 203 AA.
 XX AAG63826;
 AC AAG63826;
 XX 29-OCT-2001 (first entry)
 DT Amino acid sequence of a human zidnt5 polypeptide.
 XX Human; zidnt5; anti-angiogenic; intestinal polypeptide; wound healing;
 KW extracellular matrix interaction; tumour suppression; gamete maturation;
 KW immunologic recognition; gastrointestinal irradiation; chemotherapy;
 proteolysis; apoptosis; angiogenesis; infection; cell adhesion;
 cell fusion; cell signalling; tumour; Crohn's disease; melanoma; trauma;
 inflammatory bowel disease; food poisoning; degenerative disease;
 inflammation; fertility; gamete maturation; epithelial disorder.
 XX Homo sapiens.
 OS XX Key Location/Qualifiers
 FH Modified-site 69
 FT /note= "potential N-linked glycosylation site"
 FT Modified-site 73
 FT /note= "potential N-linked glycosylation site"
 FT Region 151..161
 FT /note= "zinc-binding motif"
 XX W0200159112-A1.
 PN 16-AUG-2001.
 XX 09-FEB-2001; 2001WO-US04198.
 XX 10-FEB-2000; 2000US-0501806.
 PR (ZYMO) ZYMOGENETICS INC.
 PA Holloway JL, Sheppard PO;
 XX WPI: 2001-522477/57.
 DR N-PSDB; AAH74759.
 XX New anti-angiogenic intestinal polypeptides, zidnt5 polypeptides, which
 PT are members of disintegrin proteases, for modulating extracellular
 PT matrix interaction, tumour suppression and wound healing -
 XX Claim 1: Page 73-74; 92pp; English.
 XX The present sequence represents a human zidnt5 polypeptide. The zidnt5
 CC polypeptide is an anti-angiogenic intestinal polypeptide. Zidnt5 is
 CC used for modulating extracellular matrix interactions. Zidnt5
 CC polypeptide is useful as a tool for identifying new family members
 CC of polypeptides. Zidnt5 polynucleotides are useful as probes or primers
 CC to clone 5' non-coding regions of zidnt5 gene. Zidnt5 polypeptides are
 CC used for tumour suppression, gamete maturation, immunologic recognition,
 CC and growth and differentiation either working in isolation or in
 CC conjunction with other molecules in colon, small intestine, fetal lung,
 CC testis and B-cells. zidnt5 polypeptides are also useful for promoting
 CC wound healing, in the treatment of disorders associated with recovery
 CC after gastrointestinal irradiation, chemotherapy or antibody use.
 CC as anti-infectives, and extracellular matrix repair and remodeling. The
 CC polypeptides are also useful for modulating proteolysis, apoptosis,
 CC angiogenesis, infection, cell adhesion, cell fusion and signalling.
 CC The polypeptides are also useful for treating tumour formation, Crohn's
 CC disease, inflammatory bowel disease, food poisoning, melanoma,
 CC degenerative diseases, disorders related to immunity, inflammation,
 CC fertility, gamete maturation, immunology, trauma and epithelial
 CC disorders.
 XX Sequence 203 AA;
 SQ Query Match 99.6%; Score 770; DB 22; Length 203;
 Best Local Similarity 99.3%; Pred. No. 9.4e-33;
 Matches 147; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAGGILHLELLVAVGPDVFOAQHEDTERVYLTNLNIGALLRDPSPSLGAQFRVHLVKWVIL 60
 DB 2 AAGGILHLELLVAVGPDVFOAQHEDTERVYLTNLNIGALLRDPSPSLGAQFRVHLVKWVIL 61
 QY 61 TEPEGAPNITANLTSSLSVCGWSQINPEDDTPGHADLVLYITRFDELDPGNRQVRG 120
 DB 62 TEPEGAPNITANLTSSLSVCGWSQINPEDDTPGHADLVLYITRFDELDPGNRQVRG 121
 QY 121 VTOLGGACSPWNSCLITDGTGFDLGVTI 148


```
XX PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX PS  
XX PS Claim 11; Page 4050; 5507pp; English.  
XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
CC antiproliferative; antiparkinsonian; nontropic; neuroprotective;  
CC osteopachic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypocensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX PS  
XX SQ Sequence 118 AA;  
  
Query Match 68.4%; Score 529; DB 21; Length 118;  
Best Local Similarity 97.1%; Pred. No. 1.7e-20;  
Matches 100; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 46 LGAQFRVHLVWVILTEPEGAPNITANLTSSLISVCGWSQTINPEDDTDFGHADLVLYIT 105  
DB 1 LGAFRHLVWVILTEPEGAPNITANLTSSLISVCGWSQTINPEDDTDFGHADLVLYIT 60  
  
QY 106 RFDELDPDGNKVRGVGTOLGACSPWWSCLITDGTGDLGVTI 148  
DB 61 RFDELDPDGNXAVRGVGTOLGACSPWWSCLITDGTGDLGVTI 103  
  
RESULT 10  
AAU05396  
ID AAU05396 standard; Protein; 26926 AA.  
XX AC AAU05396;  
XX DT 24-OCT-2001 (first entry)  
XX DE Human titin (connectin) protein sequence.  
XX KW Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;  
XX KW titin-related disease; zebrafish; heart failure; heart disease.  
XX OS Homo sapiens.  
XX PN WO200151666-A1.  
XX PD 19-JUL-2001.  
XX PF 12-JAN-2001; 2001WO-US01212.  
XX PR 12-JAN-2000; 2000US-0175787.  
XX (GENO ) GEN HOSPITAL CORP.  
XX PA Fishman MC;  
XX PI  
XX WPI; 2001-451869/48.  
DR N-PSDB; AA05390.
```

```
XX PT Determining if a subject has or is at risk of developing a  
PT titin-related disease or condition, particularly heart failures,  
PT comprises detecting the presence of a mutation in the titin gene -  
XX PS  
XX PS Disclosure; Page 57-111; 114pp; English.  
XX CC The present sequence representing human titin (also known as connectin)  
CC is described in an invention relating to a novel method for determining  
CC whether a subject has or is at risk of developing a titin-related  
CC disease or condition. The method comprises analysing a nucleic acid of  
CC a sample from the subject and detecting the presence of a mutation  
CC (e.g. the pickwick mutation in the cardiac specific exon N2B) in the  
CC titin gene, which indicates that the subject has or is at risk of  
CC developing a titin-related disease. The zebrafish which has a phenotype  
CC similar to mammalian heart failure is used as a model. The method is  
CC useful for detecting an increased likelihood of heart disease, such as  
CC heart failure, in a patient, so that appropriate intervention can be  
CC instituted before any symptoms occur. The method may also be used to  
CC facilitate determination of etiology of an existing heart condition,  
CC such as heart failure, to identify compounds that can be used to treat  
CC or prevent heart conditions, in prenatal genetic screening, e.g. to  
CC identify parents who may be carriers of a recessive titin mutation.  
CC Compounds identified using the methods may be used to treat patients  
CC that have or are at risk of developing heart disease, e.g. heart  
XX PS  
XX SQ Sequence 26926 AA;  
  
Query Match 35.2%; Score 272; DB 22; Length 26926;  
Best Local Similarity 0.4%; Pred. No. 0.45;  
Matches 101; Conservative 16; Mismatches 23; Indels 24887; Gaps 28;  
  
QY 9 ELL-----VAUGPDV----- 11  
DB 190 ELLVQEEVEVPKTKTIVSTAQISESRQTRIEKKIEAHFARSIAIVEMVIDGAAGQOL 249  
QY 12 -----VAUGPDV----- 11  
DB 250 PHKTPRIPPKPSRPTPPSIAAKAQLARQSPSPIRHSPVRHVRAPTPSPVRSP 309  
QY 12 -----VAUGPDV----- 18  
DB 310 AARISTPISRSVPLLMRKTOASTVATGPEVPPMPKQEGYVASSSEAEEMRETTLTSTO 369  
QY 19 ----- 18  
DB 370 INTEERWEGRYGVQEQVTISGAAGAAASVSASAYAAEAVATGAKEVKQADKSAAVATV 429  
QY 19 ----- 18  
DB 430 VAAVDMARVREPVISAVEOTAQRTTTTAVHIQPAQEQVRKEAEKTAATVKKVVAADKAKEQ 489  
QY 19 ----- 18  
DB 490 ELKSRTEIITTKQEQMHVTHEQIRKETKTEKTPVKVVISAAKAKEQETRISEBITKKOK 549  
QY 19 ----- 18  
DB 550 VTQEAIMKETRTVVPVKVIVATPKVKEQDLVSRGREGITTKREQVQITQEKVRKEAEKTA 609  
QY 19 ----- 18  
DB 610 LSTIAVATAKAKEQETILTRETMTATROEQIOVTHGKVDVGKKAEEAVATVAAVDQAVR 669  
QY 19 ----- 18  
DB 670 EPREPGHLEESYAQOTTLEYGYKERISAAKVAEPPORPASEPHVVPKAVKPRVIOAPSET 729  
QY 19 ----- 18  
DB 730 HIKTTDQGMHISQIKKTTDLTTLRVLHVDRKPRTASPHFTVSKISVPKTEHGYEASTA 789
```

QY 19 ----- 18
Db 790 GSAIATLQKELSATSSAQKITKSVKAPTVPKSETRVRAEPTPLPQPFADPTDTYKSEAG 849
QY 19 ----- 18
Db 850 VEVKKEGVGSIQTGVREERFEVLHGREAKVTETARPAPVEIPVTPPTLVSLGNVTVI 909
QY 19 -----FO-----AHQEDTERV- 30
Db 910 EGESVLECHISGYPGPTVWYREDYQIESSIDFQITFOSGIARLMIREAFEDSGRFTC 969
QY 31 ----- 30
Db 970 SAVNEACTVSTCYLAVOVSEEBEKKETTAVTEKFTTEEKRFVESRDVMTDLSLTEEQAG 1029
QY 31 ----- 30
Db 1030 PGEPAAPYFITKPWQKLVGGSWFGCVGGNPKPHVYWKSGVPLTTGTRYKVSYNQ 1089
QY 31 ----- 30
Db 1090 TGECKLVISMTFADDAGEYTVVVRNKHGETSASASLEEADYELLMSQOEMLYQTQVTA 1149
QY 31 ----- 30
Db 1150 FVQPEVGETAPGVYSEYEKEQEALIRKQMAKDTVVVRYVEDQEFHISSFEERLI 1209
QY 31 ----- 30
Db 1210 KEIERYIIKTTLEELLEDEGEKMAVDISESEAVESGDLRIKKNRYILEGMGVTFHCXWS 1269
QY 31 ----- 30
Db 1270 GYPLXIAWYKQKRIKHGERYQMDFLQDGRASLRIPVLPPEDEGIYTAFAASIKGNAFC 1329
QY 31 ----- 30
Db 1330 SGKLYVEPAAPLGPATYIPTLEPVSRISLSRPSVSRSPIRMSPARMSPARMSPA 1389
QY 31 ----- 30
Db 1390 RMSPGRRLEETDESQLERLYKPVFLKPVSKLEGOTARFOLKVVGRMPETFWHDQ 1449
QY 31 ----- 30
Db 1450 QIWN DYTHKVIKEDGTQSLIIVPATPSDGEWTVVAQNAGRSSISVILTVEAVEHQVK 1509
QY 31 -----LTNLNI----- 36
Db 1510 PMFVEKLVKNVNIKEGSRLEMKVRATGNPNPDIVWLKNSDIIVPHKYPKIRIEGKGEAAL 1569
QY 37 ----- 36
Db 1570 KIDSTVSQDSAWYTATAINKAGRDTRCKNVVEFAEPEPERKLIIPRGTYRAKEIAAP 1629
QY 37 ----- 36
Db 1630 ELEPLHLRYCQEWEGDLYDKKKOOKPFFKKLTSRLKRLRFGPAHFECRLTPISDPTWV 1689
QY 37 ----- 36
Db 1690 VEWLHDGKPLEANRLRMINEFGYCSLDYGVAYSRDSGIITCRATNKYGTDHTSATLIVK 1749
QY 37 -----GA----- 38
Db 1750 DEKSLVEESQLPEGRKGLORIIELEMAHEGALTGVTTDQEKQKPDIVLYPEPVRLVG 1809
QY 39 ----- 38
Db 1810 ETARFCRVGTGYPQKNNYLNQGLIRKSRFRVYDGIHYLDIVDCKSYDTGEVKVTAE 1869
QY 39 ----- 38

Db 1870 NPEGVIEHKVKLEIQOOREDFRSVLRAPEPRPEFHVHPEPGKLOFEVQKVDPRVDTTETKE 1929
QY 39 -----ELLR--- 42
Db 1930 VVKLRAERITHEKVPSESEELRSKFRRTEEGYEAITAVELSKRKDESVEELLRKT 1989
QY 43 ----- 42
Db 1990 DELLHMTKELTEEEKKALAEKGKITPTFKPKIELSPSEAPKIFERIOSQTVGOGSDA 2049
QY 43 -----DPS----- 45
Db 2050 HPRVRVCKPDECEWYKNGVKIERSDRIYWPEDNVCELVIRDVTAEDSASIMVKAIN 2109
QY 46 ----- 45
Db 2110 IAGETSSHAFLLVQAKLITFTQELQDVVAKEKDTMATFECETSEPFVKVWKYKDGMEVH 2169
QY 46 ----- 45
Db 2170 EGDKYRMHSDRKVHFLSILTITDTSDAEDYSCVLVEDENVKTTAKLIVEGAVVEFVKELQD 2229
QY 46 ----- 45
Db 2230 IEVPESYSGELECIIVSPENIEGKWHYHNDVELKSNKYTITSRRGRONLTVKDVTKEDQGE 2289
QY 46 ----- 45
Db 2290 YSFVIDGKTKCKLKMKPRPIAILQGLSDQKCEGDIVQLEVKSLESVEGWMKDGQEV 2349
QY 46 ----- 45
Db 2350 QPSDRVHIVIDKQSHMLLIEDMTKEDAGNSFTIPALGLSTSGRVSVYSDVITPLKDVN 2409
QY 46 ----- 45
Db 2410 VIEGTVKAVLECKVSPDVTSVKWYLNDEQIKPDDRVQAIKVGTKQRLVINRTHASDEGYP 2469
QY 46 ----- 45
Db 2470 KLIQVRVETNCNLSVEKIKIIRGLDLCTCTQNVVFEVLSHSGIDVLMNFKDKIKPS 2529
QY 46 ----- 45
Db 2530 SKYKIEAHGKIYKLTVLNMMKDDGKYTFYAGENMTSGKLTVAGGAIKSKPLTDOTVAESQ 2589
QY 46 ----- 45
Db 2590 EAVFECEVANPDSKGEWLROGKHLPLTNNIRSESDGHKRLIIAATKLDGGEYTYKVAT 2649
QY 46 ----- 45
Db 2650 SKTSAKLKEAVKIKKTLKNLTVTETQDAVFTVELTHPNVKGVQMIKNGVVVLESNEKYAI 2709
QY 46 -----LGAQFRVHL--VKMV-- 58
Db 2710 SVKGTIYSLRIKCAIIVDESIVYGFRLGRLGASARLHVETVKIIPKPKDVTALENATVAFE 2769
QY 59 ----- 58
Db 2770 VVSHTDTPVKWFKHSVEIKPSDKHRLVSRKVKHKLMLQNISPSDAGEYTAUVGOLECKA 2829
QY 59 ----- 58
Db 2830 KLFVETLHITKMKNIIEVETKTASFECEVSHFNVPMSMLKNGVEIEMSEKFIWVQGL 2889
QY 59 ----- 58
Db 2890 HQLIIMNTSTEDSAEYTFVCGNDVOVSATLTVTPIMITSMKLDINAEKDTITFEVTNYE 2949
QY 59 ----- 58

Db 2950 GISYKWLKNGEIKSTDKCOMRTKTLTHSLNIRNVHFGDAADYTFVAGKATSTATLYVEA 3009
QY 59 ----- 58
Db 3010 RHIEFRKHIDIKVLEKKRAMFECEVSEPDITVOMMKDOELOITDRIKIQEKYVHRL 3069
QY 59 ----- 58
Db 3070 IPSTRMSDACKYTVVAGGNVSTAKLFVEGRDVRIRSIKKEVQVIEKORAVVEFVNEDDV 3129
QY 59 ----- 58
Db 3130 DAHWYKDIEINFOVQERHKYVVERIRHMFISETROSDAGEYTFVAGRNRSSVTLYVNA 3189
QY 59 ----- 58
Db 3190 PEPPQVLQELQVTVQSGKPARFCAMISGRPOPKISWYKEQLLSTGFKCKFLHDQBYT 3249
QY 59 ----- 58
Db 3250 LLLIEAFPEDAANYTCEAKNDYGVAATTSASLSVEVPEVSPDQEMPVYPPIITPLQDTV 3309
QY 59 ----- 58
Db 3310 TSEQPARFCRVSGTDLKVSWSKDKKIKPSRFRMTQFEDTYQLEIAEAYPEDEGYT 3369
QY 59 ----- 58
Db 3370 FVANNAGQVSSSTANLSLEAPESILHERIEOIEEMEMKEPSSPFLSAEEGLHSAELQLS 3429
QY 59 ----- 58
Db 3430 KINETLELSESPPYTKFDSEKGTGPIFIKEVSNADISMGDVATLSVTVIGIPKPKI 3489
QY 59 ----- 58
Db 3490 WFNGLLTPTSADYKFVFGDDHSLIILFTKLEDEGEYTCMASNDYGTICSAYLKINSK 3549
QY 59 ----- 58
Db 3550 GEGHKOTETESAVAKSLEKLGCPGPPHFLKELPIRCAQGLPALFVTVVGEPAPTWTF 3609
QY 59 ----- 58
Db 3610 KENQOLCTSVYTIHNPNGSGTFIVNDPQREDSGLYICKAENMGESTCAEELLVLED 3669
QY 59 ----- 58
Db 3670 TDMTDTPCAKASTPEAPEDFPQPLKGPVAVELDSEOEIATFVKDTILKAALITEENQOL 3729
QY 59 ----- 58
Db 3730 SYEHIKANELSSOLPLGAQELQSILODKLTPESTREFLCINGSIHFOPLKEPSPNLQL 3789
QY 59 ----- ILTEPEG- 65
Db 3790 QIVOSQKTFSEKILMPEEPETQAVLSDEKIFPSMSIEQINSITVEPLKTLAAEPGN 3849
QY 66 ----- 65
Db 3850 YPOSSIEPPMHSYLTSAEVEVLSLKEKTVSDTNREQVTLQOEQAQSALILSQAEGHV 3909
QY 66 ----- 65
Db 3910 ESLOSPOVMSOVNYEPLVPSEHSCTEGGKILIESANPLENAGQDSAVRIEGBKSLRPL 3969
QY 66 ----- 65
Db 3970 ALEEKVLLKEHSDNVVMPDPIIESKREPVAIKKVOEQGRDLSKESLLSGIPEQR 4029
QY 66 ----- 65
Db 4030 LNLKIQICRALQAAVASEQPLFSEWLRNIEKVEAVNITQEPHRHIMCMVLTSAKSVT 4089

QY 66 ----- 65
Db 4090 EEVTTIIIEDVDQMANLKMELRDALCAIYYEIDILTAEPRIQOAKTSLQEEMDSFG 4149
QY 66 ----- 65
Db 4150 SOKVEPITEPEVESKYLJISTEEVSFNVQSRVKYLDATPVTKGVASAVVSDKQDES LK 4209
QY 66 ----- 65
Db 4210 SEEKESSESSETEEVATVKIQEAEGLIKEDGPMIHTPLVDTVSEEGDIVHLTTSITNA 4269
QY 66 ----- 65
Db 4270 KEVNYFENKLVPSDBKFKCLQDQNTYTLVIDKVNTEDHQGEYVCEALNDSKGTATSAKL 4329
QY 66 ----- ANI----- 69
Db 4330 TVVKRAAPVIKRKIBPLEVALGHLAKFTCEIOSAPNVRQWFKAGREIYESDKSIRSK 4389
QY 70 ----- 69
Db 4390 YISSLEILTQVDCGEYTCKASNEYGSCTATLTVTVPGEKKVRKLLPERKPEPKEE 4449
QY 70 ----- 69
Db 4450 VVLKSVLRKRPEEBEPKVEPKLEKVKKPAVPEPPPKPVEVEVPTVTKRERKIPETK 4509
QY 70 ----- 69
Db 4510 VPEIKPAIPALPAPKPKPEAEVTKIKPPPEPEPTPIAAPVTVPVVGKKAAPKEEA 4569
QY 70 ----- 69
Db 4570 AKPGPIKGVPKTSPPIAERRKLPGSGEKPDEAPFTYQLKAVPLKFVKEIKDIL 4629
QY 70 ----- 69
Db 4630 TESEFVGSSAIFECVSPSTAITTWMKGSNIRESPKHFIA DGKORKLHIIDVQLSDAG 4689
QY 70 ----- 69
Db 4690 EYTCVLRNLGNKKTSTAKLVBEELPVRFVKLBEETVVVKGQPLYLSCENKERVVWRK 4749
QY 70 ----- 69
Db 4750 DGKIVVEKGRIVPGVIGLMRALTINDADTDAGTYTVTVENANNLECSCKVKEVIRD 4809
QY 70 ----- 69
Db 4810 WLKPIRDQHVKPGTAIFACDIAKDTPNIKWPKGYDEIPAEPNDKTEILRDGNHLYLKI 4869
QY 70 ----- 69
Db 4870 KNAMPEDIAAYAVEIEGKRYPAKLTLGERVELLKPIEDVTIYEKESASPDABEADIP 4929
QY 70 ----- 69
Db 4930 GQWKLKCELLRPSPTCEIKAEGGKRFLLHKVKLDQAGEVLYQALNAITTAILTVEIEI 4989
QY 70 ----- 69
Db 4990 DFAVPLKDVTPERRQARPECVLTREANVIWSKGPDIKSSDKFDIADGKHLVINDS 5049
QY 70 ----- 69
Db 5050 QFDEBGVYTAEBEGKKTSAKLFVTGIRLKFMSPLEDQTVKEGTATFVCELSHEKHHVW 5109
QY 70 ----- 69
Db 5110 FKNDAKLHTSRTVLIISSEKTHKLEMKEVTLDDISQIAQVKELSSSTAQLKVLADPYFT 5169

QY	70	-----	69
Db	5170	VKLHDKTAVEKEITLKEVSKDVPVKWFKDGEIIVSPKYSIKADGLRILKIKKADLK	5229
QY	70	-----	69
Db	5230	DKGEYVDCGTDKTKANVTVEARLIEVEKPLYGVEVGTATHEIELSEDPVHGQWKLK	5289
QY	70	-----	69
Db	5290	GQPLTASPDCIEIEDGKHHILHLHNCQLGMTGEVSFOAANAKSAANLKVLEPLIFITPL	5349
QY	70	-----	69
Db	5350	SDVKVFEKAEKPECEVSREPKTFRWLKGTQEIITGDDRFELIKDGTKHSMVKSAAPEDE	5409
QY	70	-----	69
Db	5410	AKYMFEAEDKHTSGKLIIEGIRLKLPLKDVTAKEKESAVFTVELSHDNIRVKWFKNDQ	5469
QY	70	-----	69
Db	5470	RLHTRSVSMODEGKTHSITFKDLSIDDTSOIRVENMGMSSEAKLTVLEGDPYFTGKLQD	5529
QY	70	-----	69
Db	5530	YTGVEKDEVILQCEISKADAPVKWFKDGEIKPSKNAVITGDKKRMILKALKSDIGQ	5589
QY	70	-----	69
Db	5590	YTCDCGTDKTSGLDIEDREIKLVRPLHSVEMETETARFETEISEDDIHANWMLKGEAL	5649
QY	70	-----	69
Db	5650	LOTPDCEIKECKIHSLVHLNCRLDQTCGVDFQAAANVSSAHLRVKPRVIGLLRPLKDVT	5709
QY	70	-----	69
Db	5710	VTAGETATPCELSYEDIPVEWYLGKLEPSDKVPRSEGVHTLTLRDVKLEDAGEVQ	5769
QY	70	-----	69
Db	5770	LTAQDFKTHANLFPVEPPVEFTKPLEDQTVREGATAVLECEVSRENAKWKWFKNGTEILK	5829
QY	70	-----	69
Db	5830	SKKYEIVADGRVRLVHDCPTEDIKTYTCDAKFKTSCNLNVPPHVEFLPLTLQVR	5889
QY	70	-----	69
Db	5890	EKEMARFECELSRENAKVKWFKDGAIEIKGKYDIISKGAVRILVINKCLLDDEAEYSCE	5949
QY	70	-----	69
Db	5950	VRTARTSGMLTVLEEAFTKNLANIEVSETDTIKLVCEVSKPGAIEVIWYKGDEEIIETG	6009
QY	70	-----	69
Db	6010	RYEILTEGRKRLVIONAHLEDAGNYNCRPLPSSRTDGVKVVHDLAAEFISKPONLEILEG	6069
QY	70	-----	69
Db	6070	EKAEFVCSISKESFPQWKRDDKTLESQKYDVIADGKKRVLVVKDATLQDMCTYVVMVG	6129
QY	70	-----	69
Db	6130	AARAAHLTVIEKLRIVVPLKTRVKEQEVWFENCEVNTGAKAKWFRNEEAIFDSSKYI	6189
QY	70	-----	69
Db	6190	ILOKDLVTLIRDAHLDDQANVSLTNHRGENVKSAANLIVEEEDLRIVEPLKDIETW	6249
QY	70	-----	69

Db	6250	EKSVTFWCKVNLNLVTLKWTKNGBEVPFDRVSVYRDVKYKHLMTIKOCPPDEGEYIVT	6309
QY	70	-----	69
Db	6310	AGQDSVAELLIIEAPTEFVEHLEDTVTEPDPAVPSQLSREKANVKYRNGREIKEGK	6369
QY	70	-----	69
Db	6370	KYKFEKDSIHRLIILKDCRLDDECEVACGVEDKRSARLFVEEIPVEIIRPPQDILEAPG	6429
QY	70	-----	69
Db	6430	ADVFLAELNKDKVEQWLRNNMVVQGDQHOMSEGIHRLQICDIKPRDOCEYRPIAK	6489
QY	70	-----	69
Db	6490	DKEARAKLEAAAPKIKTADQDLVVDVGKPLTMVVPYDAYPKAAEAWFKENEPLSTKTID	6549
QY	70	-----	69
Db	6550	TTAEQTSFRILEAKGDKGRYKIVLQNKHGKAEFINLKVIVDPGVRNLEVTETPDGEV	6609
QY	70	-----	69
Db	6610	SLAWBEPLTDGSKIIGYVYVVERRDIKRKTWVLATDRAESCEPTVTGLQKGVLYLFRVA	6669
QY	70	-----	69
Db	6670	RNRVGTGEVETDNVPEARSKYDVPGPPLNVTITDVNRFGVSLTWEPPEYDGAIEITNV	6729
QY	70	-----	69
Db	6730	IELRDKTSIRWDTAMTVRAEDLSATVTVVGOEYSFRVRAQNRIGVGPASAATPFVKVA	6789
QY	70	-----	76
Db	6790	DPIERPSPVNLTSSTQTSVOLKWEPLKOGGSPILGVIIECEGKDNWIRCNKLV	6849
QY	77	-----	76
Db	6850	PELTYKVTGLEKNKYLRYSAENKAGVSDPSEILGPLTADDAFVEPTMOLSAFKDGLV	6909
QY	77	-----	76
Db	6910	IVPNPITILVPSTGYPRPTATWCFGDKVLETGDRVKMKTL SAYAELVISPSERSDKGIYT	6969
QY	77	-----	76
Db	6970	LKLENRVKTTISGEIDVNVVIARPSAPKELKFGDITKDSVHLTWBPDDGGSPLTGYVVEK	7029
QY	77	-----	81
Db	7030	REVSRTWTKWMDPFTDLEFTVPDLVOGKEYLFPKVCARNKCGPGEPAYVDEPVNMSTPAT	7089
QY	82	-----	81
Db	7090	VPDPENVKWRDRTANSIFLTWDPKNDGGSRIKGYIVERCPGSDKWKVACGEPVAETKM	7149
QY	82	-----	81
Db	7150	EVTGLEGKWYAVRVKTLNROGASKPSRPTEEIOAVDTQEAPEIFLDVKLLAGLTVKAGT	7209
QY	82	-----	81
Db	7210	KIELPATVTKPEPKITWTAKMILKQDKRITTIENVPKSTVTIVDSKRSDTGTVIIBAV	7269
QY	82	-----	81
Db	7270	NVCGRATAVVEVNLDPGPPAADIOTVNESCLLTWNPPRDGGSKITNVVVERRATD	7329
QY	82	-----	81

Db 7330 SEVHHKLSSTVKDTNFKATKLI PNKEYIFRVA AENMYGAGEPVQASPI TAKYQFDPGPP 7389
Qy 82 ----- 81
Db 7390 TRLEPSDITKDAVTLTWCEPDDGGSPITGYWVERLDPD TKWRCNKMPVKD TTYRVKG 7449
Qy 82 ----- 81
Db 7450 LTNKKYRFRVLAENLAGCPKSKSTEPILIKDIPDPWPPGKPTVKDVCKTSVRLNWK 7509
Qy 82 ----- 81
Db 7510 PEHGGAKIESYIEMLKTGTDEMURVAEGVPTQHLLPGLMEGOEYSFRVRAVNKAGES 7569
Qy 82 ----- 81
Db 7570 EPSEPSDVLCREKLYPPSPRMLEVINITKNTADLKWTPKXDKGSPITNYIVEKRDR 7629
Qy 82 ----- 81
Db 7630 RKGWQTVDTTVKCTVTPLTGSLYVFRVA AENAIQSDYTEIEDSVLAKOTFTTGP 7689
Qy 82 ----- 81
Db 7690 PYALAVVDVTKRHVDLKWEPKNDGGRPIQRYVIEKKERLCTR WVKAGTAGDCNFRVT 7749
Qy 82 ----- 81
Db 7750 DVEGTEVOFQVRAENAEGVHPSEPTIILSIEDPTSPSPPLDLHVTDAGRKHIAIWK 7809
Qy 82 ----- 81
Db 7810 PPEKNGGSPIIGYHVEMCPVCTEKMVRNRP IKDLKFVKEGVVPDKYVLRVAVNAI 7869
Qy 82 ----- 81
Db 7870 GVSEPSEISENVAKDPCKPTIDLETHDIIIEGEXLSIPVPPRAVPEVPTVSMHKDGE 7929
Qy 82 ----- 81
Db 7930 VKASDRLTMKNDHISAHLEVPKSVRADAGIYITILENKLGSATASINVKVIGLPGCKDI 7989
Qy 82 ----- 81
Db 7990 KASDITKSSCKLTWEPPEFDGTPILHVLERERAGRTYIPVMSGENKLSWTVKDLIPN 8049
Qy 82 ----- 81
Db 8050 GEYFRVKAVNKVGGGEYIELKNPVIAQDPKQPPDVPDVEVHNPTAEAMITWKPLYD 8109
Qy 82 ----- 81
Db 8110 GSKIMGYIIIEKIAKEERWKRCNEHLVPIITYTAKGLEEGKEYQFRVRAENAAGISEPS 8169
Qy 82 ----- 81
Db 8170 RATPPTKAVDPIDAPKVILRTSLEVKRGDEI ALDASISGSPYPTITWIKDENWIVPEIK 8229
Qy 82 ----- 81
Db 8230 KRAAPLVRRRKGEVQEEFPVLP LQRLSIDNSKKGESQLRVDRSLRDPHGLYMIKVEND 8289
Qy 82 ----- 81
Db 8290 HGIAPACTVS VLDTPGPPINFPVFEDIRKTSVLCKWEPPDDGGSEIINYTLEKKDKTKP 8349
Qy 82 ----- 81
Db 8350 DSEWIVVTSTLRHKYSVTKLIEGKEYLFRVRAENRFGPGPCVSKPLVAKDPFGPPDAP 8409
Qy 82 ----- 81
Db 8410 DKPIVEDVTSNMLVKWNEPKONGSPILGYMLEKREVNSTHWSRVNKSLLNALKANVDGL 8469

Qy 82 ----- 81
Db 8470 LEGLTYVFRVCAENAAAGPKFSPSPDKTAHDPISPPGPIPRVTDTSSTTIELEWEPPA 8529
Qy 82 ----- 81
Db 8530 FNGGGEIVGYFVDKOLVGTNKNWRCRTEKMIKVRQVTVKEIREGADYKLRVSAVNAAGEOP 8589
Qy 82 ----- 81
Db 8590 PGETQPTVAEPQEPAPVELDVSKGQIOIMAGKTLRIPAVVTGRPVTPKVTKEEGELD 8649
Qy 82 ----- 81
Db 8650 KDRVVDNVGTSKELI IKDALRKHGRVYVITATNSCGSKFAAAARVEFVDPGVPVLDLKPV 8709
Qy 82 ----- 81
Db 8710 VTNRKMLLNWSDPEDDGGSEITGFIERKDAKMHWTROPIETERSKCDITGLLEGOEYK 8769
Qy 82 ----- 81
Db 8770 FRVIAKNKFCGCPVEIGPILAVDPLGPPTS PERLTYTERQSRSTITLDWKEPRSNNGSPI 8829
Qy 82 ----- 81
Db 8830 QGVIIIEKRHRDKDPFERNVNRCLPTTSFLVENLDEHOMYEFVRKAVNEIGESEPSLPLNV 8889
Qy 82 ----- 81
Db 8890 VIQDDEVPTIKLRSLVRGDTIKVKAGEPVHIPADVTGLPMPKIEWSKNETVIEKPTDAL 8949
Qy 82 ----- 81
Db 8950 QITKEEVSERSEAKTELSIPKAVREDKGTYYVTASNRLGVSFRNVHVEVYDRPSPRNLAV 9009
Qy 82 ----- 81
Db 9010 TDIKABSCYLTWDAPLDNGGSEITHYVIDKRDASRKAEEVNTAVAKRYGIWKLIPN 9069
Qy 82 ----- 81
Db 9070 GOYEFRAVANKYGISDECKSKVIOQPYRLPGPGPKVYLARTKGSMLVSWTPPLDNG 9129
Qy 82 ----- 81
Db 9130 GSPITGYWLEKREEGSPYMSRVSRAPITKVGLKGVEFNVPRLLEGVKYOFRAMAINAAGI 9189
Qy 82 ----- 83
Db 9190 GPPSEPSDPEVAGDP IFFPPGPPSCPEVKDKTKSSI SLGWMKPPAKDGGSPIKGYIVEMQEE 9249
Qy 84 ----- 83
Db 9250 GTTDMKRVNEPDKLIITTCCECVNPNLKLRYFRVKAVNEAGESEPSD TTGTEIPATDIOE 9309
Qy 84 ----- 83
Db 9310 EPEVFDIGAQCULVCKAGSOIRIPAVIKGRPTPKSSWEFDGKAKAMKGVHDIPEDAQ 9369
Qy 84 ----- 83
Db 9370 LETAENSSVIIIECKRSHTGKYSITAKNKAQOKTANCRVKVMDVPGPKDLKVSDITRG 9429
Qy 84 ----- 83
Db 9430 SCRLSWKQPDGDDGDRIKGYVIEKRTIDGKAWTKVNPDCGTTFFVVPDLLSEQQYFRVR 9489
Qy 84 ----- 83
Db 9490 AENRFGIGPPVETIORTTARDPIYPPDPPIKLGILITKNTVHLSMKPKPKNDGGSPVTHY 9549

QY 84 ----- 83
 Db 9550 IVECLANDPTGKKEARQCNRDVEELOFTVEDLVEGGEYEFVRKAVNAAGVSKPSATV 9609
 QY 84 ----- 83
 Db 9610 GPCDCORPDMPSIDLKEFMEVEEGTNVNI VAKIKGVFPPTLTWFKAPPKKPNKEPVLY 9669
 QY 84 ----- 83
 Db 9670 DTHVNKLVVDDTCTLVI PQSRSDTGLYITITAVNNLGTASKEMRLUNVLRGPPVGPPIKF 9729
 QY 84 ----- 83
 Db 9730 ESVSADQMTLSWFPKDDGSKI TNVIEKREANRKTWVHVSSEPKECTYITIPKLLGHE 9789
 QY 84 ----- 88
 Db 9790 YVFRIMAQNKYIGIPELDSEPETARNLFSVPGAPDKPTVSSVTRNSMTVNMEPEYDGS 9849
 QY 89 ----- 88
 Db 9850 PVTGYWLEMKDTSKRKRWNRDP IKAMTLGVSYKVYTG LIEGSDYQFRVYAINAAGVPA 9909
 QY 89 ----- 88
 Db 9910 SLPSOPATARDPIAPGPPFPKVTWTKSSADLEWSPP LKDGSKVTGYIYEYKEEGKEE 9969
 QY 89 ----- 88
 Db 9970 WEKGDKVEVRGTLVVTGLKEGAFYKFRVSAVNIAGIGSGEVTDVIEKMDRLVSPDLQL 10029
 QY 89 ----- 88
 Db 10030 DASVRDRIVVHAGVIRIIAYVSGKPPPTVWNMERTLPQEA TIEHTAISSMWIKNCQ 10089
 QY 89 ----- PEDD --- 92
 Db 10090 RSHQVYSLAKNEAGERKKTIIVDLDVPGVGTFFLAHNL TNESCKLTWFSPEDDGGS 10149
 QY 93 ----- 92
 Db 10150 PITNYVIEKRESRRRAWTPVTYVTRQNATVOGLIOG KAYFRIAAENSIGMPFVETSE 10209
 QY 93 ----- 92
 Db 10210 ALVIREPITVPERPELEVKEVTKNTVTLTNPPKYDGGSEI INYVLESRLIGTEXFHKV 10269
 QY 93 ----- 92
 Db 10270 TNDNLLSRKYTVKLGEGDYEVYRVSANIVGQKPSFCTK PITCKDELAPPTLHLDPRD 10329
 QY 93 ----- 92
 Db 10330 KLIRVGEAFALTGRYSGPKPKVSNWPKDEADVLEDDRTHIKTTPATLAEKIKAKRSDS 10389
 QY 93 ----- 92
 Db 10390 GKYCVVVENSTGSRKGFQVNVVDHPGPPVGPVSFDEVTKDYMWI SNKPPPLDDGSKI TN 10449
 QY 93 ----- 92
 Db 10450 YIIKEKVGKDVMPVTSASAKTCKVKSLLEGKDYIFRIHAENLYGISDPLVSDSNKAK 10509
 QY 93 ----- 92
 Db 10510 DRFRVPDAPQPIVTEVTKDSALVTWKNKPHDGGKPI TNYLEKRETMKSRWARTKDP IH 10569
 QY 93 ----- 92
 Db 10570 PYTKFRVPDLLEGQCYEFRVSAENEIGIGDPSPKPVFAKOPTAKPSPPVNPPEADTTC 10629
 QY 93 ----- 92

Db 10630 NSVDLTWQPPRHDDGSKILGYIIVEYQKVGBEWRRAHNTPE SCPETKYKVTLGRDQGYK 10689
 QY 93 ----- TDPGH --- 97
 Db 10690 FRVLAVNAAGESDPAHVPEPVLVKDRLEPPELILDANMAREQHIKV GDTLRLSAILKVP 10749
 QY 98 ----- 97
 Db 10750 FPKVTWKEDRDAPTKARIDVTPVSGKLEIRNAAHEDGGIYSLTVENPAGSKTVSVKVLV 10809
 QY 98 ----- 97
 Db 10810 LDKPGPPRDLLEVSEIRKDCSYLTWKEPLDDGGSVITNYVVERRDVA SAQSPLSATSKK 10869
 QY 98 ----- 97
 Db 10870 SHFAKHLNEGQYLFRVAAENQYGRFPFVETPKPIKALDPLHPGPKDLHHVDVDKTEV 10929
 QY 98 ----- 97
 Db 10930 SLVMNKPRDGGSPITGVLVEYQEBEGTDWIKFKTVTNLECVVTGL OOGKTYFRVKAEN 10989
 QY 98 ----- 97
 Db 10990 IVGLGLPDTTPIECQSKLVPPSVVELDVKLEGLVVKAGTTVRFP AIIRGVPVPTAKWTT 11049
 QY 98 ----- 97
 Db 11050 DGSEIKTDEHYTVETDNFSSVLTIKNCLRRDTGEGYQITVSNAA GSKTVAVHLTVLDVPGP 11109
 QY 98 ----- 97
 Db 11110 PTGPINILDVTPHMTISWOPPKDDGGSPVINYIYERQDTRKOTM GVSSGSSKTLKIP 11169
 QY 98 ----- 97
 Db 11170 HLQKGEYVFRVRAENKIGVPPDLDTPTVAKHKFSPSPGKPVVTDITENAATVSWTL 11229
 QY 98 ----- 97
 Db 11230 PKSDGSPITGYMERREVTGKVRVNTKPIADLKFRVTGLYEGNTYEFVRFAENLAGLS 11289
 QY 98 ----- ADLV --- 101
 Db 11290 KPSSSDPIKACRPKIPGPPINPKLKD KSRETADLVMTKPLSDGSPILGYVVECOKPG 11349
 QY 102 ----- 101
 Db 11350 TAOMNRINKDELIROCAFRVPG LIEGNEYRPRIKAANIVGEGEPRELAESVIAKDILHPP 11409
 QY 102 ----- 101
 Db 11410 EVELDVTCDVTIVRGOTIRILARVKGRPEPDITWTKGKVLVREKRVDLIQDLPRVEL 11469
 QY 102 ----- 101
 Db 11470 QIKEAVRADHGKYYIISAKNSSGHAQGSAINVLD RPPCQNLKVTNVTKENCTISWENPL 11529
 QY 102 ----- 101
 Db 11530 DNGGSEITNFIVEYRKPQKQWSIVASDVTKRLIKANLLANNEYFRVCAENKVGVGPTI 11589
 QY 102 ----- 101
 Db 11590 ETKTPIALINPIDRGEPEENLHIADKGTFFVYLKWR RPDYDGGSPNLSYHVERRLKGSDD 11649
 QY 102 ----- 101
 Db 11650 WERVHKSIIKETHYMWDRVCVENQIYEFVRQTKNEGGS DWKTEEVVVKEDLOKPVLDUK 11709
 QY 102 ----- 101

Db 11710 LSGVLTUVKAGDTIRLEAGVRGKPFPEVAWTKDKDATDLTRSPRVKIDTRADSSKFSLTKA 11769
Qy 102 ----- 101
Db 11770 KRSDGGKYVVTATNTAGSFVAYATVNVLDKPGPVRLKIVDVSSDRCTVCWDPEDDGGC 11829
Qy 102 ----- 101
Db 11830 EIQNYILEKCTKRWVWSTYSATVLTGCTTVTRLIEGNEYIFRVRAENKIGTGPTTESKP 11889
Qy 102 ----- 101
Db 11890 VIAKTYDKPCRDPPEVTKVSKEEMTVMNPPEDGKSITGYFLEKKEKHSTRWVPVN 11949
Qy 102 ----- 101
Db 11950 KSAIPERRMKVQNLDPDHEYQFRVKAENEIGIGESLPSRRPVVAKDPIEPPGPTNFRVV 12009
Qy 102 ----- 101
Db 12010 DTTKHSITLGWGPVYDGGAPIIGYVVMRPKIADASPDEGWRKCNAAAQLVRKEFTVTS 12069
Qy 102 ----- 101
Db 12070 LDENQVEFRVCAQNVGIGRPAELKEAIKPKIELEPPEIDLDASMRKLVIVRAGCPRL 12129
Qy 102 ----- 101
Db 12130 FAIVRGVPAPKVTWRKVGIDNVVRKGQVDLVTWAFVIPNSTRDDSGKYSLTLLVNPAGE 12189
Qy 102 ----- 101
Db 12190 KAVFVNRVLTDPGVSVDLKVSVDVTKTSCHVSWAPPENDGGSQVTHYIVEXREADRKTWS 12249
Qy 102 ----- 101
Db 12250 TVTPEVKKTSFHVNTLVPGENEYFRVTAVNEYGPVTPDKPVLASDPLSEDPDRKLE 12309
Qy 102 ----- 101
Db 12310 ATEMTKNSATLAWLPPLRDGGAKIDGYIISYREEQPADRWTEYSVVKDLSLVVTGLKEG 12369
Qy 102 ----- 101
Db 12370 KKYFRVAARNAGVSLPREAGVYEAKQLLPPKILMPEQITIKAGKKLRIEAAHVYCKP 12429
Qy 102 ----- 101
Db 12430 HPTCKWKKEDEVTSLSLAVHKADSSILIIKDVTRKDSGYISLTAENSSGTDQKIKV 12489
Qy 102 ----- 101
Db 12490 VMDAPGPPQPFDDISDADACSLSWHIPLEDGGSNITNYIVEKCDVSRGDWVTALASV 12549
Qy 102 ----- 101
Db 12550 TKTSCRVGKLIPOQEYIFRVRAENRFGISEPLTSPKVAQPPFGVSEPKNARVTKVKNK 12609
Qy 102 ----- 101
Db 12610 CIFVADRDPDSGGSPIIIGYLIERKERNLLWVKANDTLVRSTEYPCAGLVEGLEYSFRI 12669
Qy 102 ----- 101
Db 12670 YALNAGSPPSPKTEYVTRMPVDPGPKVIDVTKSTVSLIWARPKHDSGSKIIGFV 12729
Qy 102 ----- 101
Db 12730 EACKLPDQKWRVRCNTAPHQI POEYATATGLBEKAQYQFRAIARTAVNISPPSESDPVTI 12789
Qy 102 ----- 101
Db 12790 LAENVPPRIDLSVAMKSLTLVRKAGTNVCLDATVFGKPMPTVSMKKDGTLLKPAEGIKMAM 12849

Qy 102 ----- 101
Db 12850 QRNLCLELFSVNRKDSGDYTTITAENSSGKSATIKLVLDKPGPPASVKINKMYSDRAM 12909
Qy 102 ----- 101
Db 12910 LSWEPPLEDGSEITWYIVDKRETSRPNMAQVSATVPIITSCSVSEKLIIEGHEYOFRICAEN 12969
Qy 102 ----- 101
Db 12970 KYGVGDPVFTEPAIAKPNYDPPORCDPPVISNITKDHMTVSMKPPADDGSGSPITGVLLEK 13029
Qy 102 ----- 101
Db 13030 RETQAVNMTKVRNKPITIERLTAKTLQEGTEYEFVRTAINKAGPGKPSDASKAAYARDPO 13089
Qy 102 ----- 101
Db 13090 YPPAPPAPKPYDTTRSSVLSWGPAYDGGSPIIGYLVEVKRADSDNVRCNLQNLQK 13149
Qy 102 ----- 101
Db 13150 TRFEVTGLMEDTOYQFRVAVANKIGYSDPSDVPDKHYPKDILIPPEGEHADLRKTLILR 13209
Qy 102 ----- 101
Db 13210 AGVTMLRYVPVKRPPPKITMSKPNVNLDRIGLDIKSTDFTFLRCENYKYDAGKIYL 13269
Qy 102 ----- 101
Db 13270 TLENSCKKEYTIWVKVLDTPGPPINVTVEISKDSAYVTWEPPIDGGSPIINYVVOKR 13329
Qy 102 ----- 101
Db 13330 DAERKSHSTVTTBCSKTSFRVNLBEGKSYFFRVFAENEYIGDPCGETRADKASQTPGP 13389
Qy 102 ----- 101
Db 13390 VDLKVRSVKSCSIGWKPHSDGGSRIIGYVVVDFLTEENKQVRVMKSLSLQYSAKDLT 13449
Qy 102 ----- 101
Db 13450 EGKEYTFRVSAENENGEGTPEITVVARDDVAPDDLKGLPDLCLAKENSFRLKIP 13509
Qy 102 ----- 101
Db 13510 KGKPAVSMKKGEDPLATDTRVSVESAVNTTLIVYDCQKSDAGKYITILKNVAGTKEG 13569
Qy 102 ----- 101
Db 13570 TISIKVVGKPGIPTGPIKDFDEVTAEAMTLKWAPPKDDGSEITNYILEKSDSVNNKWVTC 13629
Qy 102 ----- 101
Db 13630 ASAVOKTTFRVLRLHEGMEYTFRVSANENYGVGEGKLSEPIVARHPFDVPDAPPPIVD 13689
Qy 102 ----- 101
Db 13690 VRHDSVSLTWTDPKKTGSPITGYHLEFBEKERNLLMKRANKTPIMRDFKVTGLTEGLE 13749
Qy 102 ----- 101
Db 13750 EFRVMAINLAGVGKPSLPSEPVVALDPIDPPGKEVINITRNSVTLIWTEPKYDGGHKL 13809
Qy 102 ----- 101
Db 13810 GYIIVEKRDLPKSKMKNHNVNPECAFTVTDLVEGGKYEFIRAKNTAGAISAPSESTET 13869
Qy 102 ----- 101
Db 13870 IICKDEYEAPTIIVLDPTIKDGLTIKAGDTIVLNAISILGKPLPKSSWSKAGDIRPSDIT 13929

QY	102	-----	101
Db	13930	QITSTPTSSMLTIKVATRKAAGEYTTATNPGTKVEHVKVTVLDVPGPPGVEISNVA	13989
QY	102	-----	101
Db	13990	EKATLTWTPPLEDGGSPIKSYILEKRETSRLLTWTVSDEIQSCRHVATKLIQOGNEYIFRV	14049
QY	102	-----	101
Db	14050	SAVNHYGKEPVQSEPVKVDVRGPPGPEKPEVSNVTNTATVSMKRPVDDGGSEITGY	14109
QY	102	-----	101
Db	14110	HVERREKSLRWVRAIKTPVSDJCKVTGLQEGSTYEFVSAENRAGIGPPSEASDVLML	14169
QY	102	-----	101
Db	14170	KDAAYPGPPSPNHVDTTTKKSASLANGRPHYDGGLEITGYVVEHOKVGDEAWIKDTTGT	14229
QY	102	-LVITRF----- 	107
Db	14230	ALRITQVVPDLQTKENYFRISAINDAGVGEPAVIPDVEIVEREMAPDFELDAELRRTL	14289
QY	108	-----	107
Db	14290	VVRAGLSIRIFVPIKGRPAPEVTTKDNINLKNRANIENTESFTLLIIPECNRYDTGKVV	14349
QY	108	-----	107
Db	14350	MTIENPAGKKSGFVNVRLDTPGPVNLNRPDTDKDSVTLHWDLPLIDGGSRTITNIVEK	14409
QY	108	-----	107
Db	14410	REATRSYSTATTKCHKCTYKVTGLSEGCYFFRVMAENEYIGIETETTEPVKASEAPS	14469
QY	108	-----	107
Db	14470	PPDSLNIMDTKSTVSLAWPKPKHGGSKITGYVIEAQRKGSQDWHITTVKGLECVVRN	14529
QY	108	-----	107
Db	14530	LTEGEYTFQMAVNSAGRSAPRESRPVIVKEQTMPELDLRGIYQKLVIAKAGDNIKVE	14589
QY	108	-----	107
Db	14590	IPVLGRPKPTVMKKGDQILKQTORVNFETTATSTILNINECVRSDSGPYPLTARNIVE	14649
QY	108	-----	107
Db	14650	VGDVITIQVHDIPGPPTGPIKFDEVSDFVTFSDPPENDGGVPISNYVVMRQDSTTW	14709
QY	108	-----	107
Db	14710	VELATTVIRTYKATRLTTGLEQYFRVKAQNRVGVGPGITSAMIVANYPFKVPGPPTQ	14769
QY	108	-----	107
Db	14770	VTAVKDSMTISWHEPLSDGGSPILGYHVERKERNGILMQTVSKALVPGNIFKSSGLTDG	14829
QY	108	-----	107
Db	14830	IAYEFVIAENMAGSKSPSEPMALDPIIDPGKPVPLNITRHTVTLKWAKPEYTCGF	14889
QY	108	-----	107
Db	14890	KITSYIVKEKRDLPNGRWLKNFNSILNEFTVSGLTEDAAEFVRIAKNAAGAISSPSEP	14949
QY	108	-----	107
Db	14950	SDAITCRDDVEAPKIKVDVKFDTVILKAGFAFRLEADVSRPPTMEWSKDGELEGT	15009
QY	108	-----	107

Db	15010	KLEIKIADPSTNLVNDSTRDSCGYTLTATNPGGFAKHIFNVKVLDRPGPEGLAVTE	15069
QY	108	-----	107
Db	15070	VTSEKCVLSMFPPLDDGGAKIDHYIVOKRETSRLAMTNVASEVQVTKLVTKLKGNEYI	15129
QY	108	-----	107
Db	15130	FRVMAVNYGVEPESEPEVLAVNPYPGPPPKNPEVTTITKDSMVVCGHPDSDGSGSEI	15189
QY	108	-----	107
Db	15190	INYIVEREDKAGORWIKCNKKTLDLRYKVSGLTEGHEYEPRIMAENAGIASPSTPF	15249
QY	108	-----	107
Db	15250	YKACDTPFKPGPPGNPRVLTSSISIAMNKPIDYGSEITGYMVEIALPEDEWQIVT	15309
QY	108	-----	107
Db	15310	PPAGLKATSYTITGLTENQOEYKIRIYAMNSEGLGEPALVPGTPKAEDRMLPEIELDADL	15369
QY	108	-----	107
Db	15370	RKVVTIRACCTLRLFVPIKGRPDPEVKWARDHGESLDKASIESASSYTLIVGNVNRFD	15429
QY	108	-----	107
Db	15430	GKYILTVNSSGSKSAFVNVRLDTPGPPQDLKVKEVTKTSVTLTWDPPLDGGSKIKNY	15489
QY	108	-----	107
Db	15490	IVEKRESTRKAYSTVATNCHKTSWKVDQLQEGCSYFVRLAENEYIGLPAETAESVKAS	15549
QY	108	-----	107
Db	15550	ERPLPGKITLMDVTRNSVLSWEKPEHGGSRILGYIVEMQTKGSKWATCATVKVTEA	15609
QY	108	-----	107
Db	15610	TITGLIQEEYSFRVSAONEKGISDPROLSUPVIAKDLVIPPAPKLLFNTFTVLAGEDLK	15669
QY	108	-----	107
Db	15670	VDVPFIGRPTPAVTHKDNVPLKQTRVNAESTENNSLLTIKDACREDVGHVYVVKLTNSA	15729
QY	108	-----	107
Db	15730	GEAIETLNVILDKPGPTGPVKMDEVTAADITLSWGPPKPDGSSINNNWVEKRDJSTT	15789
QY	108	-----	107
Db	15790	TWOIVSATVARTTIKACRLKTGCEYQFRIAENRYCKSTYLNSEPTVAQYPFKVPGPPT	15849
QY	108	-----	107
Db	15850	PVVTLSRSDSMEVQNNPEISDGGSRVIGYHLERKERNLSILWVKLNKTPIPOTKFKTGLE	15909
QY	108	-----	107
Db	15910	EGVEYEFVSAENIVGIGKPSKVSECVVARDPCDPPRPEAIIVTRNSVTLOWKKPTYDG	15969
QY	108	-----	107
Db	15970	GSKITGYIVEKKELPEGRWMAKASFTNIIDTHFEVTGLVEDHRYEFVRIARNAAGVFSEPS	16029
QY	108	-----	107
Db	16030	ESTGAI TARDEVDPRI SMDPKYKDTIVVHAGESFKVDADYIGKPIPTIQWIKGDOELSN	16089
QY	108	-----	107

Db 16090 TARLEIKSTDFATSLSVKDAVRVDSGNVILKAKNVAGERSVTNVKVLDRPPEGPVWI 16149
Qy 108 ----- 107
Db 16150 SGVTAECTLAWKPPLODGGSDIINYI VERRETSLRVVTVVDANVOTLSCKVKLLEGNE 16209
Qy 108 ----- 107
Db 16210 YTFRIMAVNKYGVGEPLSEBPVAKNPVVPDAPKAPVTTVTKDSMIVVWERPASDGG 16269
Qy 108 ----- 107
Db 16270 EILGVLEKROKEGIRWTRCHKRLIGELRLRVGTGLIENHDFRVSABNAAGLSEPPS 16329
Qy 108 ----- 107
Db 16330 AYQKACDPIYKPPNNPKVIDITRSSVFLSWSKPIYDGGCEIOGYIVEKCDVNVGWTM 16389
Qy 108 -----DL 109
Db 16390 CTPPTGINKTNIIEKLEKHEYNFRICAINKAGVGEHADVPGPPIIVEEKLEAPDIDL 16449
Qy 110 EL----- 111
Db 16450 ELRKIINIRAGSLRFPVPIKGRPTPEVKMGKVGDEIRDAAIIDVTSSFTSLVLDNVNRY 16509
Qy 112 ----- 111
Db 16510 DSGKYTLFLENSSGTKSAFVTVRVLDTPSPVNLKVTEITKDSVSIWEPPLDGGKIK 16569
Qy 112 ----- 111
Db 16570 NYIVEKREATKSYAAVNTCHNSWKIDQLEGCSYFRVTAENEYIGIGLPAQTADPIK 16629
Qy 112 ----- 111
Db 16630 VAEVPPGKITVDVTRNSVLSWTKPEHDGSKIIOYIVEMQAKHSEKMSACARVSL 16689
Qy 112 ----- 111
Db 16690 QAVITNLQTGEYLFPRVAVNEKGRSDRSLAVPIVAKDLVIEPDVKPAFSSYVQVQGD 16749
Qy 112 ----- 111
Db 16750 LKIEVPIGRPKTITWTKDGLPLKQTRINVTDSLDLTLSIKETHKDDGGQYGITVAN 16809
Qy 112 ----- 111
Db 16810 VVGKTAIEIVTLKDPKPPKPVKFDVSAESITLSWNPPLYTGCCOITNYVQKRDIT 16869
Qy 112 ----- 111
Db 16870 TTWVDVWSATVARTTLKVTKLKTGTEYQFRIFAENRYGOSFALESDDPIVAQYYPKEGPP 16929
Qy 112 ----- 111
Db 16930 GTFPATAISKDSMWIQWHEPVNNGSPVIGYHLERKERNISLWTKVNTI IHDTOFKAON 16989
Qy 112 ----- 111
Db 16990 LEEGIEYFPRVAENIVGVKASKNSECYVARDPCDPPGTPEPIMVKRNEITLOWTKPVY 17049
Qy 112 ----- 111
Db 17050 DGGSMITGYIVEKRDLPDGRWMAKSTNVITQFTVSGLTEDQRYEFRVIAKNAAGATSK 17109
Qy 112 ----- 111
Db 17110 PSDSTGPIAKOEVELPRISMDPKFRDITIVNAGETFRLEADVHGKPLPTIEWLRGDK 17169
Qy 112 -----PDGMR 116
Db 17170 EESARCEIKNTOPKALLIVKDAIRIDGGQVILRASNVAGSKSPVNVKVLDRPPEGPV 17229

Qy 117 QVRGVT-----QLGG----- 126
Db 17230 QVTGVTSKSLTWSPLQDGGSDISHYVVEKRETSRLAWTVVASEVWVNSLVKTKLLEG 17289
Qy 127 ----- 126
Db 17290 NEYVFRIMAVNKYGVGEPLSAPVLMKNPFVLPGPSKLEVTNIAKDSMTVCWNRPSDG 17349
Qy 127 ----- 126
Db 17350 GSEIIGYIVEKRRSGIRWIKCNKRRIIDLRLVTVGLTEHDEYEFPRVSAENAGVGEPS 17409
Qy 127 -----ACSP----- 130
Db 17410 ATVYKACDPVFKPGPTNAHIVDTTKNSITLAWKPIYDGGSEILGYVVEICKADEEW 17469
Qy 131 ----- 130
Db 17470 QIVTPQTLRVTRFEISKLTEHOEYKIRVCALNKVGLGEATSPGTVKPEDKLEAPELDL 17529
Qy 131 -----TWS----- 133
Db 17530 DSELRKIVVRAGSARIHIPFKGRPMPEITWSEGEFTDKVQIEKGNYVTQLSIDNCD 17589
Qy 134 ----- 133
Db 17590 RNDAGKYLKLENSSGSKSAFVTVKVLDPGPQNLAKEVRKDSAPLVNPEPIIDGGAK 17649
Qy 134 ----- 133
Db 17650 VKNYVIDKRESTRKAYANVSSKCKTSFKVENLTEGAIYFRVMAENEFVGVVPEVTVDA 17709
Qy 134 ----- 133
Db 17710 VKAAEPPSPKGVTLTDVTSQTSASLWKEPEHDGGSRLGVYVVEMQPTEKMSIVAESK 17769
Qy 134 ----- 133
Db 17770 VCNVAVTGLSSGQEQYQFRKAYNEKSGSDPRVLGVPVIAKDLTIQPSLKLPFNYSIQAG 17829
Qy 134 ----- 133
Db 17830 EDLKIEIPVIGRPRPNISHWKDGEPKOTTRVNVVEATSTVLHIKEGNKDDFGKYTVTA 17889
Qy 134 ----- 133
Db 17890 TNSAGTATENLSVILEKPPGVPRFDEVSAFVVISWEPPEPAYTGCCQISNVIIVEKRD 17949
Qy 134 ----- 133
Db 17950 TTTTTHMVSATVARTTIKITLKTGTEYQFRIFAENRYGKSAPLDSKAVIVQYYPKEPG 18009
Qy 134 ----- 133
Db 18010 PPGTFVTSISKQMLVQMHPEVNDGGTKIIGYHLEKKNISLWKLNTKPIQDTKPKT 18069
Qy 134 ----- 133
Db 18070 TGLDEGLEYPKVAENIVGIGKPSKVSCECFVARDPCDPPGRPEAIIVITRNNVTLKWKPK 18129
Qy 134 ----- 133
Db 18130 AYDGSKITGYIVEKKDLPDGRWMAKSTNVLETEFTVSGLVEDQRYEFRVIARNAAGNF 18189
Qy 134 ----- 133
Db 18190 SEPSDSSGAITARDEIDAPNASLDPKYKDVIVVHAGETVLEADIRGKPIPDVWVSKDGK 18249
Qy 134 ----- 133
Db 18250 ELEETAARMEIKSTIQKTTLVKDCIRTDGGQYILKLSNVGKTSKIPITVKVLDPRGSP 18309

QY 134 ----- 133
 Db 18310 GPLKVTGTAECYLAWNPPLODGGANISHYIIIEKRETSRLSWTQVSTEVOALNYKVTKL 18369
 QY 134 ----- 133
 Db 18370 LPGNEYIFRMAVNKYGIGEPLESGPVTACNPKPKPGPPSTPEVSAITKDSMVVTWARPV 18429
 QY 134 ----- 133
 Db 18430 DDGGTEIEGYILEKRDKEGVRWTKCNKTKTLDLRLRVTLGLTEGHSYEFRAAENAAGVGE 18489
 QY 134 ----- 133
 Db 18490 PSEPSVFRACDALYPPGPSNBKVTDTSRSSVSLAWSKPIYDGGAPVKGYVVEKEAAA 18549
 QY 134 ----- 133
 Db 18550 DEWTTCTPTGLOGKQFTVKLKENTENYFRICAINSEGVGEPATLPGSVVAQERIEPPE 18609
 QY 134 ----- 133
 Db 18610 IELDADLRKVVLRASATLRLFTVIKGRPEPEVKWEKAEGLTDRAQIEVTSFTMLVID 18669
 QY 134 ----- 133
 Db 18670 NVTRFDSGRYNLTLENNSGSKTAFVNVVLDSPSAPVNLTIREVKKDSVTLSEPPPLIDG 18729
 QY 134 ----- 133
 Db 18730 GAKITNVIIEKRETRTKAYATITNNCTKTTFRIENLOEGSYFYFRVLASNEYGIGLPAET 18789
 QY 134 ----- 133
 Db 18790 TBPVKVSEPLPGRVTLVDVTRNTATIKWEKPEDSGSKITGYVVMQTKGSEKMWSTCT 18849
 QY 134 ----- 133
 Db 18850 QVKLEATISGLTAGBEYFRVAANVEKGRSDRQLGVPVIARDIEIKPSVELPFTFNV 18909
 QY 134 ----- 133
 Db 18910 KAREQLKIDVPFKRQATVNNRKGOTLKETTRVNVSSKTVTSLIKEASKEDVGTYE 18969
 QY 134 ----- 133
 Db 18970 LCVNSAGSITVPITIIVLDRPGPGPIRIDEVSCDSITISWNPPEYDGGCQISNIVEK 19029
 QY 134 ----- 133
 Db 19030 KETTSTTHIVSQAVARTSIKIVRLTTGSEYQFRVCAENRYGKSSYSESSAVVAEYFPSP 19089
 QY 134 ----- 133
 Db 19090 PGPPGTPKVHATKSTMLVTMQVPNDGGSRVIGYHLEYKERSILLWSKANKILIADTV 19149
 QY 134 ----- 133
 Db 19150 KVSGLDEGLMYEYRYAENIAGIGKSKCEPVPARDPCDPGQPEVTNITRKSLSUKWS 19209
 QY 134 ----- 133
 Db 19210 KPHYDGAKITGYIVERRELDPGRWLKCNYNIOETFEVTELTEDORYEFERVFARNAAD 19269
 QY 134 ----- 133
 Db 19270 SVSEPESTGPIIVKDDVEPRVMVDVKFRDVIIVVKGAEVLKINADIAGRPLPVISNAKD 19329
 QY 134 ----- 133
 Db 19330 GIEIERARTEIISTDNHTLLTVKDCIRRDGTQYVTLKNVAGTRSVAVNCKVLDKPGPP 19389
 QY 134 ----- 133

Db 19390 AGPLEINGLTAECSLSGWRPOEDGGADIDYIHRKKRETSHLAWTICEGELQMTCKVTK 19449
 QY 134 ----- 133
 Db 19450 LLKONEYIFRVTGVNKYGVGEPLSVAKALDPFTVPSPPTSLEITSVTKESMTLCHSRP 19509
 QY 134 ----- 133
 Db 19510 ESDGSEISGYIIERREKNSLRWRVNVKPYVDLVRKSTGLGECEGEYERYVAENAGLS 19569
 QY 134 ----- 133
 Db 19570 LPSETSPILRAEDPVFLPSPSKPKIVDSGKTTITIAWVKPLFDGAPITGYTVEYKSD 19629
 QY 134 ----- 133
 Db 19630 DTDWKTISQSLRGTEYTIISGLTTGAEYFVRVKSUNKVGCASDPSSDDPOIAKEREPEPLF 19689
 QY 134 ----- 133
 Db 19690 DIDSEMRKTLIVKAGASFTMTVPFRGRPVNVLMKSKPDTDLRTRAYVDTTDSRTSLTIEN 19749
 QY 134 ----- 133
 Db 19750 ANRNDGKYTLTIONVLASAASLTLLVVKVLDTPGPPNTITVODVTKESAVLSWDVPENDGG 19809
 QY 134 ----- 133
 Db 19810 APVKYHIEKREASKAWWSVTNNCNRLSYKVTNLOEGAIYYFRVSGENEFQVGIPAETK 19869
 QY 134 ----- 133
 Db 19870 EGVKITEKPSPEKLGVTISKDSVSLTWLKPHEHGGSRIVHYVVEALEKGOKNWKCAV 19929
 QY 134 ----- 133
 Db 19930 AKSTHHVVSGLRENSSEYFVRFAENQAGLSDPRELLLPVLIKEQLEPPEIDMKNPSTHV 19989
 QY 134 ----- 133
 Db 19990 YVRAGSNLKVDPISGKPLPKVTLSDRGVPLKATMRFNTEITAENLTINLKESVTADAGR 20049
 QY 134 ----- 133
 Db 20050 YEITANSSGTTKAFINIVLDRPGPTGPVVISDITEBSVTLKWEPPKYDGSQVTNYI 20109
 QY 134 ----- 133
 Db 20110 LLKRETSTAVMTEVSATVARTMMKVMKLTGTGEYOFRIKAENRFGISDHQACVTVKLP 20169
 QY 134 ----- 133
 Db 20170 YTTGPPSPWNTVNTRESITVGWHEPVSNGSAVVGYHLEMKDRNSILLWOKANKLVIRT 20229
 QY 134 ----- 133
 Db 20230 THPKVTISAGLIYEFVRVAENAGVGKPSHPSEPVLAIDACEPPRNVIRITDISKNSVL 20289
 QY 134 ----- 133
 Db 20290 SMOOPAFDGGSKITGYIVERRDLDPGRWTKASFNTVETQFTISGLTONSQYEFVRFARN 20349
 QY 134 ----- 133
 Db 20350 AVGSIENPSEVGPITCIDSYGGPVIDLPLEYEVVKYBAGTSVKLRAGISGKPAPTEW 20409
 QY 134 ----- 133
 Db 20410 YKDDKELOTNALVCVENTTDLASILIKADRLNSGCEYELKLRNANASATIRVQILDKP 20469
 QY 134 -----CLIT 137
 |||

Db 20470 GPPGPIEFKTVTAETKILLWRPPADGGAKITHYIVKEKRETSRVVWSMVSEHLEECIT 20529
Qy 138 ----- 137
Db 20530 TTKIIKGNEYIFRVRANKYGIGEPLESDSVAKNAFVTPGPGIPEVTKITKNSMTVW 20589
Qy 138 ----- 137
Db 20590 SRPIADGSDISGYPLEKXRDKSLGWFVLKETIRDTQKVTGLTENSQYQYRCVAVNAA 20649
Qy 138 ----- 137
Db 20650 GQPFSEPFYKAADPIDPPGPAKIRIADSTKSSITLGWSKPYDGSVAVTGYVWEIR 20709
Qy 138 ----- 137
Db 20710 QGEBEEMTVSTKGEVRTTEYVWSNLKPGVNYFRVSAVNCAGOGPIEMNEPVOAKDIL 20769
Qy 138 ----- 137
Db 20770 EAPELDLVALRTSVIAKAGEDVQLIPFKGRPPPTVTRKDEKVLGSDARYSIENTDSS 20829
Qy 138 ----- 137
Db 20830 SLLTIPOVTRNDTGKYLITLTIENGVEPKSSTVSVKVLDTPAACQKLVKHVSRGVTLLW 20889
Qy 138 ----- 137
Db 20890 DPPLIDGSPINNVIEKRDATKRTWSVSHKCSSTSKLIDLSEKTPFFFRVLAENIG 20949
Qy 138 ----- 137
Db 20950 IGECETTEPVKAAEVPAPIRDLSMKDSTKTSVILSWTKPDPDGGSVITEYVVERKKGKE 21009
Qy 138 ----- 137
Db 21010 QTWSHAGISKTEIEVSQLEQSVLEFRVFAKNEKGLSDPVTIGTITVKELIITPEVDLS 21069
Qy 138 ----- 137
Db 21070 DIPGAQVTVRIGHNVHLEPYKGPSPISWLKQGLPLKESEFVRFSKTENKITLSIKNA 21129
Qy 138 ----- 137
Db 21130 KKEHGKTVILDNAVACRIAPITVITLGPSPKPGPIRPFDEIKADSVILSHDVPEDNGG 21189
Qy 138 ----- 137
Db 21190 GEITCYSIEKRETSQTNMKWCSSVARTTKVPNLVKDAEQFRVAENRYGVSOPLVSS 21249
Qy 138 ----- 137
Db 21250 IIVAKHOFRIPOPPGKPVINYVTSQMSLTWDAPVYDGGSEVTGFHVEKKERNILWQV 21309
Qy 138 ----- 137
Db 21310 NTSPISGREYRATGLVEGLDYQFRVYAENSACLSPSPDSKFTLAVSPVDPGTPDYIDV 21369
Qy 138 ----- 137
Db 21370 TRETITLKNPPLRDGSKIVGYSIEKQGNRWVRCNFTDVSECOYTVTGLSPGDRYEF 21429
Qy 138 ----- 137
Db 21430 RIIRNAVGTISPPSQSGIIMTRDENVPPIVEFGPEYFDGLIIKSGESLRIKALVQGRP 21489
Qy 138 ----- 137
Db 21490 VPRVTWFKDGEIEKRMMEITNVLGSTSLFVRDATRDHRGVYTVBEAKNASAKAEIKV 21549
Qy 138 ----- 137
Db 21550 KYQDTPGVVGPIRFTNITGEKQTLWMDAPLNDGAPITHYIIEKRETSRLAWALIEDKC 21609

Qy 138 ----- 137
Db 21610 EAQSYTAIKLINGEQFRVSAVNKEFGVGRPLSDPVAQIQYVTPDAPGIPBSNITGN 21669
Qy 138 ----- 137
Db 21670 SITLTWARPESDGGSEIQYILERREKKSTRVVKVISKRPISETRFKVTGLTEGNEYEFH 21729
Qy 138 ----- 137
Db 21730 VMAENAAVGVPASGISRLIKCREPVPNPPGPTVVKVTDTSKTTVSLWSKPVFDGGMELI 21789
Qy 138 ----- 137
Db 21790 GYIEMCKTDLGDWHKNAEACVKTRYVTDLQAGEEYKFRVSAINGAGKGDSEVGTGI 21849
Qy 138 ----- 137
Db 21850 KAVDRLTAPELDIDANFKQTHVVRAGASIRLFIAYQGRPTPTAVWSKPDNSLSLRADIHT 21909
Qy 138 ----- 137
Db 21910 TDSFSTLTVCNCRNDAGKYTLTVENNNGSKSITFTVKVLDTPGPPGPIITFKOVTRGSAT 21969
Qy 138 ----- 137
Db 21970 LMWDAPLLDGGARIHHYVVEKREASRSQVISEKCTRIQIFKYNDLAEGVPPYFRVSAVN 22029
Qy 138 ----- 137
Db 22030 EYGVGEPEMPEPIVATEQPAPRRLLDVDTSKSSAVLAWLKPDDHGGSRITGYLLEMRO 22089
Qy 138 ----- 137
Db 22090 KGSDLWVEAGHTKOLTFTVERLVEKTEYEPFRVAKNDAGYSEPREAPSSVIIKEPQIEPT 22149
Qy 138 ----- 137
Db 22150 ADLTGINTQLITCKAGSPFTIDVPISGRPAPKVTWLEEMRLKETDRVSITTTKDRITLT 22209
Qy 138 ----- 137
Db 22210 VKDSMRDGSRYFLTLENTAGVKTFTVVVVVIGRPGVPTGPIEVSSVSAESCVLWSGEPK 22269
Qy 138 ----- 137
Db 22270 DGGTEITNIVIEKRESGTTAMQLVNSSVKRTQIKVTHLTKYMEYSFRVSENRFVSKP 22329
Qy 138 ----- 137
Db 22330 LESAPIIAEHFPVPPSPAPTREPVYHVSAVNSIRWEEPYHDGSKIIGYMWVEKKERNITL 22389
Qy 138 ----- 137
Db 22390 WVKENKVPCLCNKYKVTGLVEGLEYPQRTVALNAAVSKASEASRPIMAQNPFVDAPGRPE 22449
Qy 138 ----- 137
Db 22450 VTDVTRSTVSLIMSAPAYDGGSKVGVYIIBERKPVSEVGDGRWLKCNITYIVSDNFTVTAL 22509
Qy 138 ----- 137
Db 22510 SEGDTYFRVLAKNAAGVISKGSSESTGPTCRDEYAPPKAELDARLHGLDVTIRAGSDLV 22569
Qy 138 ----- 137
Db 22570 LDAAVGKPEKPIIWTGDKKELDLCEKVSLOYTQGRATAVIKFCDRSDSGKYTLTVKNAS 22629
Qy 138 ----- 137
Db 22630 GTKAVSVNVKVLDSPPGCGKLTVSRVTOEKCTLAWSLPQEDGGAEITHYIVERRETSRLN 22689

Qy	138	-----	137
Db	22690	WVIVEGECPTLSVYVTRLKNNKEYIFRVRANKYGGVPESEPIVARNSFTIPSPGP	22749
Qy	138	-----	137
Db	22750	EEVGTGKEHIIQWTKPESDGGNEISNYLVDRKESLRWTRVNDYVYVDTRLKVTSLM	22809
Qy	138	-----	137
Db	22810	EGCDYQFRVTAVNAGNSEPSESNFISCREPSYTPGPPSAPRVVDTTKHSISLAWTKPM	22869
Qy	138	-----	137
Db	22870	YDGGTOIVGVLEMOEKDOWYVHTNATIRNTEFTVPLKMGOKYSFRVAANVKGMS	22929
Qy	138	-----	137
Db	22930	EYSESIAETPVERIEIPDLELADDLKKTVIRAGASLRMLVSVGRPPPVITWSKQIGD	22989
Qy	138	-----	137
Db	22990	LASRAIIDTTESYLLIVDKVNRVDAGKYTIEAENOSGKSATVLVKVYDTGPGCPVKV	23049
Qy	138	-----	137
Db	23050	KEVSRDSVTITWEIPTIDGAPINNVIKREAAARAFKTVTKCKTLVYRISGLVEGTM	23109
Qy	138	-----	137
Db	23110	HYFVLPENIYGIGECETSDAVLSEVPLPAKLEVVDTKSTVTLANEKPLYDGSRL	23169
Qy	138	-----	137
Db	23170	TGYVLEACKAGTERMMKWVLKPTVLEHTVTSLEGEQYLFIRIAQNEKGVSEPREVTA	23229
Qy	138	-----	137
Db	23230	VTQDLRLPTIDLSLTPQKTIHVPAGRPVELVPIAGRPPPAASWFFAGSKLRESERV	23289
Qy	138	-----EDTG-----	141
Db	23290	VETHKVAKLTIRETTIRDTGEYTLKQVNTGTSETIKVILDKPGPTGPIKIDEIDA	23349
Qy	142	-----	141
Db	23350	TSITISWEPPELDGGAPLSGYVWEQDAHRPGWLPVSESVTRSTFKFTRLTEGNEYVFRV	23409
Qy	142	-----	141
Db	23410	AATNRFIGSYLQSEVIECRSSIRIPGPPETLQIFDVSRDGMTLTWYPPEDDGGSQVTGY	23469
Qy	142	-----	141
Db	23470	IVERKEVRADRWVRVNVKVPVMTVRYRSTGLTEGLEYEHRVTAINARGSGKPSRSPKIVA	23529
Qy	142	-----	141
Db	23530	MDPIAPGPKQNPVRTDTRTSVSLAWSVPEDEGSKVTGYLIEQVKVDQHEWTKCNTTP	23589
Qy	142	-----	141
Db	23590	TKIREYTLTHLPQGAEYFRVLACNAGGPEAEVPGTVKVTMELVYDPYELDERYQEGI	23649
Qy	142	-----	141
Db	23650	FVROGGVIRLTIPIKGPFPICKWTKEGODISKRAMIATSETHTELVIKEADRGDSGYD	23709
Qy	142	-----	141
Db	23710	LVLENKCGKAVYIKVRVIGSPNPEGPLEYDDIQVRVRSWRPPADGGADILGYILE	23769
Qy	142	-----	141

Db	23770	RREVPKAAWVTIDSRVRGTSLSLVKGLKENVEYHFRVSAENQFGISKPLKSEEPVTPKTPL	23829
Qy	142	-----	141
Db	23830	NPPEPPSNPPEVLVDTKSSVLSWSRPKDDGSGRVGTGYIERKETSTDKWRHKNKTQITT	23889
Qy	142	-----	141
Db	23890	TMYTVTGLVPDAEQFRIIAQNVDGLSETSPASEPVVCKDPFKPSQPOGELEILSISKDS	23949
Qy	142	-----	141
Db	23950	VTLOWKEPCDGGKEILCYWVEYRQSGDSAMKSNKERIKDKQFTIOGALLEATEYEFVRF	24009
Qy	142	-----	141
Db	24010	AENETGLSRPRTAMSIKTKLTSGEAPGIRKEMKDVTKLGEAAQLSCQIVGRPLPDIKM	24069
Qy	142	-----	141
Db	24070	YRFKELIOSRKYKMSSDGRTHLTVMTEEOEDEGVYTCIATNEVEGETSSKLLQATP	24129
Qy	142	-----	141
Db	24130	QPHPGYPLKEYGAVGSTLRHLVHMYIGRPVPMTFMHGOKLLQNSENITIENTEHTHL	24189
Qy	142	-----	141
Db	24190	VMKNVORKTHAGKYKVQLSNVFGTVDAILDVEIQDKPDKPTGPIVIEALLKNSAVISMKP	24249
Qy	142	-----	141
Db	24250	PADGGSMITNVYVEKEGAEWQLVSSAISVTTCRIVNLTENAGYFVRVSAQNTFGI	24309
Qy	142	-----	141
Db	24310	SDPLEVSVVIIKSPFEPGAPKPTITAVTKDSCVWANKPPASDGGAKIRNYYLEKREK	24369
Qy	142	-----	141
Db	24370	KONKIVSTTEIRETVFSVKNLIEGLEYPFKVCNLCGESESEISEPITPKSDVPITQ	24429
Qy	142	-----	141
Db	24430	APHFKEELRNLRVQSNATLVCKVTHPKPIVKWYRQGEIIADGLKYRIOEFKGGYHQ	24489
Qy	142	-----	141
Db	24490	LIIASVTDDATVYQVRATNOGGSVSGTASLEVEVPKIHLPKTLGCMGHALRGEVVS	24549
Qy	142	-----	141
Db	24550	IKIPFSGKPDPIVWQKQODLIDNNCHYQVIVTRSFVSLVFPNGVERKDAGFYVVCANR	24609
Qy	142	-----	141
Db	24610	FGIDQKTVELDVADVPDPRGVKVSARSRSVNLWTPEPASDGGSKITNYIVEKCAATAE	24669
Qy	142	-----	141
Db	24670	RWLRVGOARETRYTVINLFGKTSYQPRVIAENKFGLSKSPSESEPTITKEDKTRAMNYDE	24729
Qy	142	-----	141
Db	24730	EVDETREVSMTKASHSTKELYEKYMIADLGRGEFGIVHRCVETSSKTYMAKFKVKG	24789
Qy	142	-----	141
Db	24790	TDQVLVKKEISILNIAHRNHLHESFESMEELVMIFISGLDIFERINTSAFELNER	24849
Qy	142	-----FDL-----	144

Db 24850 EIVSYHVCEALQFLSHNIGHDIRPENIIYQTRRSSTIKIIEFQCARQLKPGDNFRL 24909
QY 145 ----- 144

Db 24910 LFTAPEYAPEVHQHDVYVSTATDMWSLGTLYVYVLLSGINPFLAETNOQIIENIMNAEYTF 24969
QY 145 ----- 144

Db 24970 DEAFKEISIEAMDVDRLLVKERSMTASEALQHPMLKQKIERVSTKVRTLKHRYY 25029
QY 145 ----- 144

Db 25030 HTLIKDLNMVSAARISCGAIRSQKSVAKVKVASIEIGPVSGQIMHAVGEGGHVK 25089
QY 145 ----- 144

Db 25090 YVCKIENYDOSTQVTWYFGVROLENSEKYEITYEDGVAILYVVKDITKLDGTYRCKVND 25149
QY 145 ----- 144

Db 25150 YGEDSSVAELFVKGVREVDYCRRTMKIKRRTDTMRLLERPEFTPLYNKNTAYYGEN 25209
QY 145 ---GVTI 148
|||

Db 25210 VRFGVTI 25216

RESULT 11
AAR44929
ID AAR44929 standard; Protein; 15281 AA.
XX
AC AAR44929;
XX
DT 08-JUL-1994 (first entry)
XX
DE T. niveum Cyclosporin synthetase.
XX
KW Enzyme; cyclosporin; synthetase-like activity; Tolypocladium niveum;
KW T. inflatum GMS; biosynthesis; vector; cyclosporin synthetase.
XX
OS Tolypocladium niveum.
XX
PN EP578616-A.
XX
PD 12-JAN-1994.
XX
PF 05-JUL-1993; 93EP-0810474.
XX
PR 09-JUL-1992; 92AT-0001403.
PR 08-MAR-1993; 93AT-0000437.
PR 29-APR-1993; 93CH-0001310.
PR 04-MAY-1993; 93CH-0001375.
XX
PA (SANO) SANDOZ LTD.
PA (SANO) SANDOZ PATENT GMBH.
PA (SANO) SANDOZ-ERFINDUNGEN VERW GES MBH.
PI Leitner E, Schneider E, Schoergendorfer K, Weber G;
XX
XX WPI; 1994-010432/02.
XX N-PSDB; AAQ54386.
XX
PT Isolated DNA sequence - which codes for enzyme having cyclosporin
PT synthetase like activity
XX
XX Claim 1; Page 41-84; 93pp; English.
XX
CC This sequence represents an enzyme which has cyclosporin synthetase-
CC like activity. This sequence was isolated from Tolypocladium niveum
CC (formerly known as T. inflatum GMS). This enzyme catalyses the
CC peptide biosynthesis of cyclosporins and structurally related
CC molecules. This sequence may be used for the production of
CC cyclosporin by transforming a vector containing this sequence in

CC to a recombinant host. This allows effective production of anti-
CC biotic cyclosporin or its derivatives.

XX
SQ Sequence 15281 AA;
Query Match 34.2%; Score 264; DB 15; Length 15281;
Best Local Similarity 0.8%; Pred. No. 0.3;
Matches 102; Conservative 24; Mismatches 21; Indels 12475; Gaps 32;

QY 1 AAGGIL----- 6
|||

Db 2098 AAGHVLEIGTGTGMVLFNLGQAGLSYIGLEPSQSAVFVNKAQTFPGLEGKAOVHVGT 2157
QY 7 ----- 6

Db 2158 AMDTGRLSALSPDLIVINSVAQVFPSPREYLAEVVEALVRIPGVRRIFPFGDMRTYATHKDF 2217
QY 7 -----HLELL----- 11
|||

Db 2218 LVARAVHTNGSKVTRSKVOQEVARLEELLDVDPAFFTSLSKESIEIEHVEILLPKNM 2277
QY 12 ----- 11

Db 2278 KVNELSSYRYGAVLHNRHNNQNSRSHKINAESWIDFASSQMDRQGLARLLKENKDAE 2337
QY 12 ----- 11

Db 2338 STAVFNIPYSKTIIVERHIAKSLADDHGDGDDTHSSIDGVAWISAAREKASQCPSLDVHDLV 2397
QY 12 ----- 11

Db 2398 OLAEDAGFRVSWARQSONGALDVFFHHFQPTENESRALVDFTDYKGOQARSUTNRP 2457
QY 12 ----- 11

Db 2458 LORVESRRIEAQVREQLQVLLPAYMIPARIVVLQNMPLNTSGKVDKELTLRAKVTAART 2517
QY 12 ----- 11

Db 2518 PSSELVAPRDSIEAIIKEFKDVLGVEGIDTNFNFVNGHSLATKLAARLSROLNAQIA 2577
QY 12 ----- 11

Db 2578 VKDIFDRPVIADLAATIOQDTTEHNPIPTSYTGPVEQSAOGRWLFLDQLNVGATWYLM 2637
QY 12 ----- 11

Db 2638 PFAVRLRGLVVSALAAALLALEERHETLRTTFIEQEGIMQVIHPFAPKELRVIDVSGE 2697
QY 12 ----- 11

Db 2698 EESTIQKILEKEQTTPFNLAEPGFRLLKGTGDESHILSTMHHAISDGWSVDIFQOEI 2757
QY 12 ----- 11

Db 2758 GQFYSAILRGHDPLAQIAPLSIQYRDFATWQRIQVQAEHRRQLAYWTKOLADNKPALL 2817
QY 12 ----- 11

Db 2818 TDFKRPPLMSGRAGEIPVVVDGLIYEKLODFCIRQVTAFTVLLAAPRAAHYMTGTEDA 2877
QY 12 ----- 11

Db 2878 TICTPIANRRPELEGLIGFFVNTQCMRITVDVEDSFETLVHQVRETTLAAHANQDVFFE 2937
QY 12 ----- 11

Db 2938 QIVSNILPGSSDTSRNPVLQMFALHSQONLKGVRLEGIEEIIISIAETTRFDIEFHLQ 2997
QY 12 ----- 11

Db 2998 EAERLNGSIVYAADLVFPETIQSVITIFOGILOKGLGCEPDMPVASMALDGLSLSRSTGL 3057

QY 12 -----VAVGPDV----- 18
Db 3058 LHPQOTDPCDASVQIFKQAVNPDVIAVRDESTRLSYADLRKSDQVACWLSRRGIA 3117
QY 19 ----- 18
Db 3118 PETFAILAPRSCETIVAILGVLKANLAYPLDVNVPASRLAAILSEVSGSMLVLVGAET 3177
QY 19 ----- 18
Db 3178 PIPEGMAEATIRITEILADAKTDDINGLAASOPTAASLAYVIFTSGSTRPKGVMEHR 3237
QY 19 ----- 18
Db 3238 GIVRLTKOTNITSKLPESPHMAHJSNLAFDASVWEVFTLLNGGTLVCIDYFTLLESTAL 3297
QY 19 ----- 18
Db 3298 EKVFQORVNVALLPPALLKQCLDNSPALVKLTLSVLYIGDRLDASDAKARGLVQTQAF 3357
QY 19 ----- 18
Db 3358 NAYGPTENTVMSITFYPIAEDPPINGVPIGHAVSNSGAFVMDQNOQITPPGAMGELIVTGD 3417
QY 19 ----- 18
Db 3418 GLARGYTTSSLNTGRFINVDIGEQRVAYRTGDRVRYRPKDLOIEFFGRIDHQVKIRGHR 3477
QY 19 ----- 18
Db 3478 IEPAEVEYALLSHDLVDAAVTHSOENQDLEMVGVFAARVADVREDESSNQVQEWQTHF 3537
QY 19 ----- 18
Db 3538 DSITAYADITIDQOSLGRDFMSWTSMDGSLIKKSQWELDDTMRSLDSDPPGHVLEV 3597
QY 19 ----- 18
Db 3598 GTGTGMVLFNLREGGLOSYGLEPSFSATAFVNKAAKSPFPLEDRIRVEVGTATDIDL 3657
QY 19 ----- 18
Db 3658 GDDLHAGLVVNSVAQYFFSQDYLAQLVRDLTKVPGVERIFFGDMRSHAINRDLVARAV 3717
QY 19 ----- 18
Db 3718 HALGDKATKAEIQREVVRMESEDELLVDPAPFTSLTQVENIKHVEILLPKMRATNELS 3777
QY 19 ----- 18
Db 3778 SYRYAAVLHVNDLAKPAHKVSPGAWVDFAATKMDRDALIRLLRGTKISDHIAIANIPNSK 3837
QY 19 ----- 18
Db 3838 TIVERTICESVYDLGGDAKSDNRVSWLSAARNAVKVASLSAIDLVDIAQEAAGRVEIS 3897
QY 19 ----- 18
Db 3898 CARWSONGALDAVFHHLGPSPOSSHVIDFLTDHQRPPEALTNHPLHRAQRRRVERQI 3957
QY 19 ----- 18
Db 3958 RERLOTLLPAYMIPAQIMVLDKLPNANGKVDRKQLTQRAQTVPKAKQVAPVAPRTEIE 4017
QY 19 ----- 18
Db 4018 RVLQBFSDVLGVLDIGIMENFFDLGGHSLMATKLAARISRRLETHSVKIEIFDHPRVCDL 4077
QY 19 ----- 18
Db 4078 VLIVOGSAPHDPVSTKYTGVPQSFQGRLWFLDQNFQATWYMLPLAVRLRGAMNVH 4137
QY 19 ----- 18

Db 4138 ALTAALLALERRHELLRTTFYEQNGVMQKVPVWVTTETRIIDLNSGDDGYLPTLKKQET 4197
QY 19 ----- 18
Db 4198 APFHLETEPGWRVALLRLGPDYILSVVMHHIISDGSVDPVLFOELQOFYSTAVKGHDPL 4257
QY 19 ----- 18
Db 4258 SOTPLPIHYRDPALMOKKPTQSEHERQLOYWVEQLVDSAPAEALLTDLPRPSILSGOAG 4317
QY 19 ----- 18
Db 4318 EMSVTIEGALYKNLEEFRCVRHRTVSFVLLAALRAAHVRLTGSSEDATIGTBIANRNPRL 4377
QY 19 ----- 18
Db 4378 EQIIIGFFVNTOCIRITVNEDETFESLVQOVRSTATAAFAHODVPFEKIVSTVLLPGSRDAS 4437
QY 19 ----- 18
Db 4438 RNPLVQLMFAVHSOKNLGELKENAHSEVVPTEITTRFDELFHLFOODDKLEGSILYSTD 4497
QY 19 ----- 18
Db 4498 LFEAVSVQSLLSVFQEILRRGLNGPDVPISLPLQDGIQDLQROGLLDVQKTEYPRDSSV 4557
QY 19 ----- 18
Db 4558 VDVFEQVINSIPDSIALIHGSEKLSYAQLDRESORVARWLHRHSFSBSTLIAVLAPRSC 4617
QY 19 ----- 18
Db 4618 TIIAPLGLKANLAYPLDVKAPAARIDAIIVSSLPGNKLIILGANVTPPKLOEAAIDFVP 4677
QY 19 ----- 18
Db 4678 IRDTFTLLTDLQDPTIERPSAQSLAYAMFTSGTGRPKGMVQHRNIVRLVKNSNV 4737
QY 19 -----FOAHO---- 23
Db 4738 AKQPAARIAHISNLAFDASSWEIYAPLLNGGAVICADYFTTIDPQALOETFOEHEIRGA 4797
QY 24 ----- 23
Db 4798 MLPPSLLKQCLVQAPDMISRDLILPAAGDRFSSVDALQAQRLVGSVFNAYGPTENTILS 4857
QY 24 ----- 23
Db 4858 TIYNVAENDSFVNGVPIGSAVNSGAYIMDKNQQLVPAGVMGELVVITGDCARGYMDPKL 4917
QY 24 ----- 23
Db 4918 DADRFIOLTVNGSEQVRAIRTGDRVRYRPKOFQIEFFGRMDQOIKIRGRIEPAEVEQNF 4977
QY 24 ----- 23
Db 4978 LNDGFVEDVAIVIRTPENQEPENVAFTAKGDNARSAREEATTQIEGWEAHEGGAYANIE 5037
QY 24 -----EDT----- 26
Db 5038 EIESEALGYDFMGWTSMYDGTEDIDKEMREMLNDTMRSLLDGKPGAVLEVGTGTGMINF 5097
QY 27 -----ERYV----- 30
Db 5098 NLGRSQGLERYIGLEPAPSAAEFVNNAKSPPLAGRAEVHVGTAADVGTLOGLTSDMAV 5157
QY 31 ----- 30
Db 5158 INSVAQYFPTPEYLAETIKSLVQVPMKRIYLGDMRSMWNRDPAAARAAYSADNASK 5217
QY 31 ----- 30

Db 5218 RVQKMELEEKEELLVDPAFFALTASOLOQRIQHVEILPKRMKATNELSSRYAAVLH 5277
Qy 31 ----- 30
Db 5278 ISDEPLPIYKIDPEAWINFGSRLTRELAQVLKENENAESVAISNIPYKTVVERHIVR 5337
Qy 31 ----- 30
Db 5338 SLOEDANAPEESMGSDMISAVRTAQOCHTILSASLDPIAEDAGFRVEVSWARQHSQH 5397
Qy 31 -----LTV----- 33
|||
Db 5398 GALDAVFHLKPADESRVLKIFPTDHQGRPLKSLTNQPLLPQAQSRRAELLIREGLQTL 5457
Qy 34 ----- 33
Db 5458 PPTWIPSOITLIDRMPLNANGKVDRLARRAKITQKSKPVEDIVPPRNSVEATVCKGFT 5517
Qy 34 ----- 33
Db 5518 DVLGVEGITDNFFNLGGHSLMATKLAARLGRQLNTRISVRDVFDPQPVADLAAVIQRS 5577
Qy 34 -----LNIG----- 38
|||
Db 5578 APHEPIKPADYTGVPVQSFQAGRLWFLDQNLVNGATWYMLPLGIRLHGLSRVDALATAISA 5637
Qy 39 ----- 38
Db 5638 LEORHEPLRTTFHEEDGVGVQVQDHRPKDLRIIDLSTQPKDAYLAVLKHQETTLFDLAT 5697
Qy 39 ----- 38
Db 5698 EPGWRVALIRLGBEEHILSIVMHHIISDCHSVEVLPDEMHRFYSSALRQDPMEQILPLP 5757
Qy 39 ----- 38
Db 5758 IQYRDFAAQWKTBEQVAEHQRLDYMTHELADSTPAELLTDLPRPSILSGRANELPLTIE 5817
Qy 39 ----- 38
Db 5818 GRLHDKLRAFRCRVHOATPFVILLAAARAHYRLTGAEDATLGTPIANRNREPELENMIGFF 5877
Qy 39 ----- 38
Db 5878 VNTQCMRIAIEENDNFESLVRVRSTATSAFANQDVPFESIVSSLLPGSRDASRNPLVQV 5937
Qy 39 ----- 38
Db 5938 ILAVHSQODLGLTLEGLRDEAVDSATSTRFDVEFHLFEHADRLSGSVLYAKELFKLRTI 5997
Qy 39 ----- 38
Db 5998 ESVSVFLETLRALDQPLTFLAVLPLTDGVGGEIASKGLLDVPRDTPYPRDANIVEVQFH 6057
Qy 39 ----- 38
Db 6058 VRATPDAIVKDATSILTYAQDQSDRLAIWLSRRHMPETLVGLAPRSCETIIAMFG 6117
Qy 39 ----- 38
Db 6118 IMKANLAYPLDINSAPAARLSILSAVDGNKLVLLGSGVTAPEQENPEVAVGIOELAG 6177
Qy 39 ----- 38
Db 6178 TGLDKTQGSNARPSATSLAYVIFTSGTGPKGMVMEHRSVTRLAKPSNVISKLPQGARV 6237
Qy 39 ----- 38
Db 6238 AHLANIAFDASIEWEIAATLLNGATLVCLDYHTVLDCTRLKEVFERESITVVTMLPALLKQ 6297
Qy 39 ----- 38
Db 6298 CVAEIPETLAHLDDLTYTGGDRVGGHDAMRARSVKIGMFGSGYGPTENTVISTIVEVDADE 6357

Qy 39 ----- 38
Db 6358 MFVNGVPIGKTVNSGAYVMDRNOQLVPSPGVVGGELVVTGDLGARGYTDPSLNKNRFIYT 6417
Qy 39 ----- 38
Db 6418 VNGESIRAYRTGDRVRYRPHDLQIEFFGRMDQOVKIRGHRIEPGEVESALLSHNSVQDAA 6477
Qy 39 ----- 38
Db 6478 VVICAPADODSAGEMVAFVAARNTEDETOBEAEAVDOVQGWETHFETAAYSEVKDIROSE 6537
Qy 39 ----- 38
Db 6538 VGNDFMGWTSMDGSEIDKIDMHEWLNDTMRMILDAREPHVLEICTGTGTMVMFNLAACP 6597
Qy 39 ----- 38
Db 6598 GLQGYVGPFPKSAAQFVNDAASPALKDGRSIVHVGTATDINKAGPIQPRLLVINSVA 6657
Qy 39 ----- 38
Db 6658 QYFPTPEYLFVRVVEALVQIPSVVERIVFGDMRTNAINRDFVASRALHTLQOKANKRLVRQM 6717
Qy 39 -----ELLRDP----- 45
|||
Db 6718 IYELEANEHEELTDPAFFTSLRGLGEKIKHVEILLPKTKMATNELSKYRYAAVLHVGRSR 6777
Qy 46 ----- 45
Db 6778 EQSTIHQVSPNAWIDFAADGLDROTLINLLKEHKDAGTVAIGNIPYSKTIIVERFVNKSLS 6837
Qy 46 ----- 45
Db 6838 EDDMEGQNSLDCSAAWAAVRMAAOSCPSLDAMDVKEIAQEAQYQVEVSWARQWSONGAL 6897
Qy 46 -----LGAQ----- 49
|||
Db 6898 DAIFHHFEPKEGARTLIEFPTDYEGRNVTLTNRPLNSIQSRRLGTQIREKLTLLPPV 6957
Qy 50 ----- 49
Db 6958 MIPSRIWLDQMPVNNNGKIDRKELVRRAIIVAPKPRSAATRVAPRNEIBAILRDEFEDVL 7017
Qy 50 ----- 49
Db 7018 GTEVSVLDNFFDLGGHSLMATKLAARVSRRLDAHISIKDVFQOPVLADLAASIQRESAPH 7077
Qy 50 ----- 49
Db 7078 EPIQRPYTGPAEQSPAQGLWFLDQNLGATWYLMPLAIRIGQLRVAALSAAFLALER 7137
Qy 50 ----- 49
Db 7138 RHETLRTTFEESDGVGVQIVGEARNSDLRVDVSTGDDGEYLEVLRRQTVPPDLSSEPG 7197
Qy 50 FRVHLVK----- 56
:|:|:|
Db 7198 WRVCLVKTGEEDHVLIVMHHIYDGSVDILRGELGQFYSAALRGQDPLLHANPLPIQY 7257
Qy 57 ----- 56
Db 7258 RDEAANQREAKQVEEHORQLGYWSKQLVDSTPAELLTDLPRPSILSGRAGSVDTVIEGSV 7317
Qy 57 -----MWILT----- 61
:|:|
Db 7318 YGALQSFCTRSTVTFEWLLTVFRIAHFRLTAVDDATIGTPIANRNREPELETIVGCFVNT 7377
Qy 62 ----- 61
Db 7378 QCMRISIADDDNFEGLRVQRNVATAAYANQOVPPFERRIVSALVPGSRNTRSNPLVQLMFA 7437

Qy 62 ----- 61
Db 7438 VQSVEDYQVRLGLESVMFGEASTRDFMEFHLVPGQKLTGSLYSSDLFEOGTQNF 7497
Qy 62 ----- 61
Db 7498 VDIQECULRSVLDQPLTPISVLPFSNAISNLESLLLEMPSTDYPRDRTVVDLFRQAAI 7557
Qy 62 ----- 61
Db 7558 CPDSIAVKDSSQLTYAQLDEQSDRAAWLHERHMPAESLVGLSPRSCETIIAYFGIMK 7617
Qy 62 ----- 61
Db 7618 ANLAYPLDVYAPDARLAAILDVGERLLLLGAGVPOPGIQI PRLSTAYIAEALSHATT 7677
Qy 62 ----- 65
Db 7678 VDVTSI POPSATSLAYVIFTSGTGKPGVMIEHRGIVRLVRDTNVNVPESGSALPVSH 7737
Qy 66 ----- 65
Db 7738 FSNLAWDAATWEIYTAVLNGTVCIDRDTMLDIAALNSTFRKENVRAAFPTPAFLKQCL 7797
Qy 66 --APNITANL----- 73
Db 7798 AETPELVANLEILHTAGORLDPGDANLAGKTAKGGINVLGHTENTAYSTFYPPVGEETF 7857
Qy 74 ----- 73
Db 7858 VNGVPGRGISNHAVIDRHOKLPAGVMGELLTGDGVARGYDTSALNKDRFVYIDIN 7917
Qy 74 ----- 73
Db 7918 GKSTWSYRTGKARYPRDQLEFFGRMDQMKIRGVRIEPEGEVELTLLDHKSVLAAVW 7977
Qy 74 ----- 73
Db 7978 VRRPPNGDEPIAFITIDAEDDVQTHKAIYKHLQILPAYMIPSHLVLDQMPVTDNGKV 8037
Qy 74 ----- 73
Db 8038 DRKDALARAOQVQKRSTAAVPRDEVEAVLCEEYSNLEVEVGITDGFGLGGHSLLA 8097
Qy 74 ----- 73
Db 8098 TKLAARLSQLNTRVSKDVFDQPI LADLADIIRRGSHRHPIPATPYTGPVEQSFAGR 8157
Qy 74 ----- 73
Db 8158 LWFLEQLNLGASWYLMPPFAIRMRGPLOTKALAVALNALVHRHEALRTTFEDHDGVGVQVI 8217
Qy 74 ----- 73
Db 8218 QPKSSODLRIIDLSDAVDDTAYLAALKREOTTAFDLTSEPGWRVSLRLGLGDDDDYILSIW 8277
Qy 74 ----- 73
Db 8278 HHIISDGWTVDVLRQELGOFYSAAIRGOEPLSOAKSLPIQYRDFAVWQROENIQEAKQ 8337
Qy 74 ----- 73
Db 8338 LKYWSQOLADSTPCFELTDLPRPSILSGEADAVPMVIDGTIVYQLLTDFCRTHQVTSFVSL 8397
Qy 74 ----- 73
Db 8398 LAAFRTAHYRLTGLDATVGTPIANRNRPELEGLIGFFVNTQCMRAISETETFESLVQO 8457
Qy 74 ----- 73
Db 8458 VRLTTTEAFANODVPFEQIVSTLLPGSRDTSRNPLVQVMFALQSQODLGRIOLEGMTDEA 8517
Qy 74 ----- 73

Db 8518 LETPLSTRLDLEVHLFOEVGKLSGLLYSTDLFEVETIRGIVDVFLFI LRRGLEOPKQRL 8577
Qy 74 ----- 73
Db 8578 MAMPIITDGIITKLRDQGLLTVAKPAYPRESVIDLFRQOVAAPDAIYVWSSSTLYADL 8637
Qy 74 ----- 73
Db 8638 DGOSNKLAWLHLCORNMAPETLVAVFAPRSCLTIVAFGLVKANLAYLPLDVNAPARIAEA 8697
Qy 74 ----- 73
Db 8698 ILSAVPGHKLVLQAHGPELGLTMADETQIDEALASSSSGDHEQTHASGPTATSLAYV 8757
Qy 74 ----- 73
Db 8758 MFTSGTGKPGKGMIDHRSIIRLVKNSDVVATLPTPVRMANVSNLAFDISVQEIYLTALLN 8817
Qy 74 ----- 73
Db 8818 GGTLVCLDYLTLDDSKILYNVFEAQVNAAMFTPVLLKQCLGNMPAISRLSVLFNVGDR 8877
Qy 74 ----- 73
Db 8878 LDAHDAVASGLIQDAVYNAYGPTENGMSQSTMYKVDVNEPFGVPGRSITNSGAYVMD 8937
Qy 74 ----- 73
Db 8938 GNQOLVSPGVNGEIVTGDGLARGYDSDALDEDRFVHVTTIDGSENIKAYRTGDRVYRPX 8997
Qy 74 ----- 73
Db 8998 DFEIEFFGRMDQVKIRGHRIEPAEVEHALLGHDLVHDAVAVLRKANOBPEMIAFITSQ 9057
Qy 74 ----- 73
Db 9058 EDEIEQHSNKVQGVGHEFVDSRYADIKDLDTSTFGHDFLGTSMYDGVDPVNEKME 9117
Qy 74 ----- 78
Db 9118 WLDETTASLLDNRRPPGHILEIGAGTGMLSNLKGVDGLQKVGLDPAASAIFVNEAVKS 9177
Qy 79 ----- 78
Db 9178 LPSLAGKARVLVGTALDIGSLDKNEIQPELVWINSVAQYFPTSEYLIKVKVAVVEVPSVK 9237
Qy 79 ----- 78
Db 9238 RVFFGDIRSOALNRDFLAARAVRALGDNASKEQIREKIAELEBSEELLVQPAFFVSLRS 9297
Qy 79 ----- 78
Db 9298 QLPNIKHEVLPKLMKATNELSSYRYAAVLHISHNEEQLLIQDIDPTAWVDFAATOKDS 9357
Qy 79 -----SVC 81
Db 9358 QGLRNLLOQRDDVMIAVGNIPYSKTI VERHIMNSLDQDHVNSLDGTSWISDARSAAAIC 9417
Qy 82 ----- 81
Db 9418 TSFDAPALTQAKBEGFRVELSWARQSRONGALDAVPHRLATDANCERSRVLVHPTDQ 9477
Qy 82 ----- 81
Db 9478 GRQLRTLNRPLQRAQSRRIESQVFEALQTPALPAYMIPSRIVLPPQMPNTANGKVDKQOL 9537
Qy 82 ----- 81
Db 9538 ARRAQVAVKAVSARVAPRNDTEIVLCEEYADILGTEVGITDNFDMGGHSLMATKAA 9597
Qy 82 ----- 81

Db 9598 RLSRLDTRVTVKEVDFKPVLDLAASIEOGSTPHLPPIASSVYSGPVEQSYAQGRMLWLD 9657
Qy 82 ----- 81
Db 9658 QFNLNATYHWSLAWRLGLPLNDALDVALRALBORHETLRTTTEAQKDIGVQVVHEAGM 9717
Qy 82 ----- 81
Db 9718 KRLKVLDSLSDKNEKHEMAVLENEQMRPFTLASRPGWKGHRLARLGPTEYILSLVMHMFSD 9777
Qy 82 GWS ----- 84
|||
Db 9778 GWSVDILRQELGQFYSAALGRDPLSQVKPLPIQYRDPAAWKEAAQVAEHEROLAYWEN 9837
Qy 85 ----- 84
Db 9838 QLADSTPGELLTDPPRPOFLSGKAGVIPVTIEGVPYKLLKFSKERQVTLFSLVLLTAPRA 9897
Qy 85 ----- 84
Db 9898 THFRLTGAEDATIGTPIANRNRPELEHIIIGFFVNTQCMRLLLDTGSTPESLVQHVRSVAT 9957
Qy 85 ----- 84
Db 9958 DAYSQDIPFERIVSALLPGSDASRSPLIQMFALHSQDPLGNITLREGLERLPTSA 10017
Qy 85 ----- 88
|||
Db 10018 TRFDMFHLFOEPNKLSGSILFADELFPQETINSVVTVFQILRGLDQPOVSISTWMLPT 10077
Qy 89 ----- 88
Db 10078 DGLIDLEKLGLEIESSNPPRDYSVVDVFRQOVAANPNAPAVDSETSWSYTSLDQKSEQ 10137
Qy 89 ----- 90
|||
Db 10138 IAAWLHAQGLRPESLICYMAPRSPETIVSLFGILKAGAYLPLDVNSPAARIQIPILSEVE 10197
Qy 91 ----- 90
Db 10198 GKRLVLLGSGIDMPOSMDVETARIQDILNTKVERSDPMSRPSATSLAVVIFTSGTG 10257
Qy 91 ----- 90
Db 10258 RPKGMIEHRNLRVLKQSNVTSQPLQDLRMAHISNLAFAFSIWEIFTAILNGGALICID 10317
Qy 91 ----- 90
Db 10318 YFTLLDSQALRTTFEKARVNATLFPALLKECLNHAPTLFEDLKVLYIGDRLDATDAK 10377
Qy 91 ----- 90
Db 10378 IQALVKGTYNAYGPTENTVMSTIYRLTDGESYANGVPIGNAVSSGAYIMDQKORLVPP 10437
Qy 91 ----- 90
Db 10438 GVMGELVSGDGLARGYNSTLNADRFVDIVINDOKARAYRGDRTRYPKDGSTIEFFGR 10497
Qy 91 ----- 90
Db 10498 MDQQVKIRGHRVEPAEVEQAMLGNAIHDAVVVQAVDQETEMIGFVSMASDRFESEGE 10557
Qy 91 ----- 93
|||
Db 10558 EITNQVQEMEDHFESTAYAGIEAIDQATLGRDFTSWTSWYNGNLIDKAEMEELDDTMOQ 10617
Qy 94 ----- 93
Db 10618 LLDKEDARPCAEIGTGTGNVLENLPKNDCLESYVVGIEPSRSALFVDKAAQDFPGLOGKT 10677
Qy 94 ----- 93
Db 10678 QILVGTAEIDIKLVKDFPDVWVINSVAQYFPRSRYLVQIASELIHMTSVKTIFFGDMRSW 10737

Qy 94 ----- 95
|||
Db 10738 ATNRDFLVSRLYTLGDKATKQOIROEVARLENEDELLVDPAPFTSLTSQWPKVKHVE 10797
Qy 96 ----- 95
Db 10798 ILPKRMTSNELSSYRYAAVLHICRDGGRNRYGRRVHSVEENAWIDFASGMDRHALVQ 10857
Qy 96 ----- 95
Db 10858 MLDERDAKTVAIGNIPHSNTINERHTTSLDTEGEGIAQDSLDSAWQSAQKAMAARCP 10917
Qy 96 ----- 95
Db 10918 CLSVTELVEIGAAGFRVEVSWARQSRQSHGALDVVFHLEDDRVGRVLINFPDTERLPP 10977
Qy 96 ----- 95
Db 10978 STGLTSRPLORIQRREFESQIREQLOTLPPYVMVPSRIVVLERMPLNANSKVDKELARK 11037
Qy 96 ----- 97
|||
Db 11038 ARTLOTIKPSATRVAPRNDIEAVLCEFOAVLGVTVGVMDNFFELGCHSLMATKLAARLS 11097
Qy 98 ----- 101
|||
Db 11098 RRLDTRVSVKDIQFNQITLQDLADVQVOTGSAPHEAIPSTPYSGPVEQSFQGRMLWLDQLN 11157
Qy 102 ----- 101
Db 11158 LNASWYHMLASRLRGLRIEALQSALATIEARHESLRTTFEODGVVPOIVRAARNKOL 11217
Qy 102 ----- 101
Db 11218 RIIDVSGTEYLAALKQEODAAFDLTAEPGRVALLRLCPDDHVLISVMHHIISDCMSV 11277
Qy 102 ----- 103
|||
Db 11278 DILRQELGQLYSNASSQAPLPIQYRDFAIWQKDSQIAHOKQLNWKRLVNSKPABL 11337
Qy 104 ----- 103
Db 11338 LADTRPKALSGDADVIPIEIDDQVYQNLRSFCRARHVSFVALLAAFRAAHRLTGAED 11397
Qy 104 ----- 103
Db 11398 ATIGSPIANRNRPELEGLICGFVNTQCLRIPVKSEDTFTLVKQARETATEAQDNQDVPF 11457
Qy 104 ----- 111
|||
Db 11458 ERIVSSMVASSRDTSRNPLVQVMFAVHSQHDGLGNIRLEGVEGKPVSMASSTRFDEMHFL 11517
Qy 112 ----- 111
|||
Db 11518 EDOGLGQNVVFSKDLFESETIRSVAVFOETLRRGLANPHANLPLTDLGSLSLSLC 11577
Qy 112 ----- 113
|||
Db 11578 LQVNPDPYPRDASVIDVFREQVASIPKSIADASSOLTYTELDERSSOLATWLRQVTV 11637
Qy 114 ----- 118
|||
Db 11638 PEELGVGLAPRSCETIIAFLGIKANLAYLPLDVNAPAGRIETILSLPGRNLLIGASDT 11697
Qy 119 ----- 119
|||
Db 11698 QAVKLHANSVRFRTRISDALVESGPPTEELSTRPTAOSLAYVMFTSGTGVKGMVMEHR 11757
Qy 120 GVT ----- 122
|||
Db 11758 GITRLVKNVNVAKQAAAAIAHLSNIAFDASSWEIYAPLLNGTGVVCIIDYTTIDIKAL 11817

QY 123 ----- 122
Db 11818 EAVFKQHHRGAMLPALLKQCLVSAPTWISSLEILFAAGDRLLSSQDAILARRAVSGVY 11877
QY 123 ----- 122
Db 11878 NAYGPTENTVLSITHIGENAEFNGVPIGNVSNAGFVMDQNOQLVSAGVIGELVWTG 11937
QY 123 ----- 122
Db 11938 DGLARGYTDKLRVDRFIYITLDGNRVRAYTGDVRVHRPKQGIIEFFQMDQOIKIRGH 11997
QY 123 ----- 122
Db 11998 RIEPAEVEQALARDPAISDAVITQLTDEEPELVAFFSLKGNANGTNGVGSDOEKID 12057
QY 123 ----- 122
Db 12058 GDEQHALLMENKIRHNLQALLPTYMIPSRIIHVDQLPVNANGKIDRNLAVRAQAATPTS 12117
QY 123 ----- 122
Db 12118 SVSTYVAPRNDIETIIEKEFADILSVRVGITDNFFDLGGHSLIATKLAARLSRLDTRVS 12177
QY 123 ----- 122
Db 12178 VRDVFDPVVGQAASIQQGSTPHEAIPALSHSGPVQQSFAOGRLWFLDRFNLAAYIM 12237
QY 123 ----- 122
Db 12238 PFGVRLRGPLRVDAQTALRALREHERLLRTTTEEQDQGMQIVHSPMRDICVWDISGA 12297
QY 123 ----- 122
Db 12298 NEDLAKLKEEQAPFNLSDEVAMRVALFKAGENHILSIVMHHIISDGWSVDIFQOELAQ 12357
QY 123 ----- 122
Db 12358 FYSVAVRGHDPLSQVRPLPIHYRDFAVWQODKQAVHESQLQYWIEQLADSTPAIILSD 12417
QY 123 ----- 122
Db 12418 FNRPEVLSGEAGTVPIVEDEVYKLSLFCRNHQVTSFVLLAAFRVAHYRLTGAEDATI 12477
QY 123 ----- 122
Db 12478 GTPIANRNRPELEDLIGFFVNTQCMRIALBEHDNFLSVVRRVRSTAASAFENQDVPPERL 12537
QY 123 ----- 122
Db 12538 VSALLPGSRDASRNPLVOLMFVVHSQRNLKQLEGLEEPTPYTATTFRDVEFHLFEQD 12597
QY 123 ----- 122
Db 12598 KGLAGNVFAADLFEAATIRSVVEVPHEILRRGLDQPDIAISTMPLVDGLAALNSRLPA 12657
QY 123 ----- 122
Db 12658 VEDIEPDFATEASVVDVFQVWNPDALAVTDTSTKLTVAELDQOSDHVAWLSKQXLP 12717
QY 123 ----- 122
Db 12718 AESIVVVLA PRSSETIVACIGILKANLAYLPMDNSVPEARQAILSEIPGEKFVLLGAGV 12777
QY 123 ----- 122
Db 12778 PIPDNKTADVWVFI SDIVASKTDKSYSGTPRPSASSLAYVIFTSGTGRPKGMVVEHRG 12837
QY 123 ----- 122
Db 12838 VISLVKQNASRIPQSLRMAHVSNLAFDASVWEI FTTLLNGTLPFCISYFTVLDSKALSAA 12897
QY 123 ----- 122

Db 12898 FSDHRINITLLPPALLKQCLADAPSVLSSLESYIGGDRLDGADATKVKDLVKGKAYNAY 12957
QY 123 ----- 122
Db 12958 OPTENSVNSTIYTIETHEFTFANGVPFGTSLGPKSKAYIMDQOQLVPAGVMGELVVGDL 13017
QY 123 ----- 122
Db 13018 ARGYTDPSLNTGRFIHITIDGKQVQAYRTGDRVYRPDYQIEFFGRLDQOIKIRGHRIE 13077
QY 123 ----- 122
Db 13078 PAEVEQALLSDSSINDAVVSAQNKEGLEMVGYITTOAAQSVDKKEASNKVQEWAEHAFDS 13137
QY 123 ----- 122
Db 13138 TAYANIGGIDRDALGQDFLSWTSMYDGLIPREEMOEHLNDTMRSLLDNQPPGKVEIGT 13197
QY 123 ----- 122
Db 13198 GTGMVLFNLKVEGLQSVAGLEPSRSVTAWYNKALETFFPSLAGSARVHVGTAEIDISSIDG 13257
QY 123 ----- 122
Db 13258 LRSDLWINSVAQYPPSREYLAELTANLIRLPGVKRIFFGDMRTYATNKDFLVARAVHTL 13317
QY 123 ----- 122
Db 13318 GSNASKAMVROQVAKLEDBELLVDPAFTSLSDQFPDEIKHVEILPKRMAANTHELSSY 13377
QY 123 ----- 122
Db 13378 RYAAVHYGGHMPNGEDEDKQWAVKDINPKAWVDFAGTMRDQALLQLLQDRQRQDDVV 13437
QY 123 ----- 122
Db 13438 AVSNIPYKTIMERHLSQSDDEDDGTSAVDGTAMISRTQRAKECPALSADVADLIEIGK 13497
QY 123 ----- 122
Db 13498 IGFEVSAWQHSORGGDLDAVFRHFPFPRHSGHVMFPFTEHKGRSSSLTNRLPHLLO 13557
QY 123 ----- 122
Db 13558 SRRLEAKVRERLQSLPPYMI PSRITLLDQMLTNSGVKDRKKLARQARVIPRASAATLD 13617
QY 123 ----- 126
Db 13618 FVAPRTEI EVLCEEF TDL LGVKV GITDNFFELGCHSLLATKLSARLSRRQDAGITVKOV 13677
QY 127 ----- 126
Db 13678 FDQPVADLAASILQSSSRHSRISPLPYEGPVEQSAOGRLWFLDQFNIDALWYLI PPAL 13737
QY 127 ----- 126
Db 13738 RMRGPLOVDALAAALVALEERHESLRTTTEBRDGVGIQVQPLRTTKDIRIIDVSGMRD 13797
QY 127 ----- 132
Db 13798 DAYLEPQKEOQTFFDLASEPGMRVALLKLGKDDHILSIVMHHIISDGWSTEVLORELQO 13857
QY 133 ----- 132
Db 13858 FYLAASKGAPLSQVAPLPIQYRDFAVWQREBQVESAQRQLDYWKQLADSSPAELLAD 13917
QY 133 ----- 132
Db 13918 YTRPNVLSGEAGSVFVINDSVYKSLVFCRSRQVTTFTLLAAPRAAHYRMTGSDATI 13977
QY 133 ----- 136
| :
SCLT

Db 13978 GTPIANRRBELENLIGCFVNTQCMRITIGDDETFESLVQOVRSTTATATENQDVPPERI 14037
Qy 137 -----TEDTGFDL----- 144
Db 14038 VSTLSAGSRDTSRNPVLVQLFAVHSQQGLGRQLDGVVDEPVLSTVSTRFDLEPHAFQEA 14097
Qy 145 ----- 144
Db 14098 DRLNGSYMFATDLFPQETIGFVAVVEVLQGLEQFQSPATMPLAEGIAQLRDAGALQ 14157
Qy 145 ----- 144
Db 14158 MPKSDYPNALSVDVFOQQAASPSVAVTDSKLYAEULDRSLDQAASVLRROQLPAE 14217
Qy 145 ----- 144
Db 14218 TMVAVLAPRSCETIIAFLAILKANLAYMPLDVNTSPARMEAIISVPGRRLLVSGVRH 14277
Qy 145 ----- 144
Db 14278 ADINVPNAKMLISDTVTGTDAIGTPEPLVVRPSATSLAYVIFTSGTGPKGVMEVHRA 14337
Qy 145 ----- 144
Db 14338 IMRLVKDSNVVTHMPPATRMMAHVNTNIAPDVSLEFEMCATLLNGGLVCIDYTLTLDSTMLR 14397
Qy 145 ----- 144
Db 14398 ETPREQVRAAIFPPALLRQCLVNMPPDAIGMLEAVYVAGDRFHSRDARATQALAGPRVYN 14457
Qy 145 ----- 144
Db 14458 AYGPTENAILSTIYNIDKHPVNGVPIGSVNSGAYVMDRNOQLPPGVMGELVVTGE 14517
Qy 145 ----- 144
Db 14518 GVARGYTASLDTRFVTVTIDGQRQRAYRTGDRVVRPKGFQIEFFGRLDQQAKIRGHR 14577
Qy 145 ----- 144
Db 14578 VELGEVEHALLSENSVTDAVVLRTMBEEDPQLVAFVTTDHEYRSGSSNEEDPYATQAA 14637
Qy 145 ----- 144
Db 14638 GDMRKLRLSLPYVMPVSRVTILRQMLNANGVKDRKDLARRAQMTPTASSGPGVHVAPR 14697
Qy 145 -----GVT 147
Db 14698 NETEAAICDEFETILGVKVGIT 14719

RESULT 12
AAB41379
ID AAB41379 standard; Protein; 1784 AA.
XX
AC AAB41379;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1143 polypeptide sequence SEQ ID NO:2286.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antitaxial; antibacterial; antifungal; antitumetic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disease; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX Homo sapiens.
XX WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 01-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR N-PSDB; AAC75588.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 1678-1682; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antitumetic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 1784 AA;

Query Match 30.8%; Score 238; DB 21; Length 1784;
Best Local Similarity 5.8%; Pred. No. 0.038;
Matches 70; Conservative 25; Mismatches 51; Indels 1067; Gaps 13;
Qy 1 AAGGILHLEL-----LVAVG----- 15
Db 57 AGGSFLSYELWPRALRKRDVSVRRDAPAFYELQYRGRELNLNTANOHLLAPGPFVSETRR 116
Qy 16 -----PD-----VFOAHQED----- 25
Db 117 RGLGRAHIRAHTPACHLLGEVDQPELEGLAAISACDGLKGVFOLSNEDYFIEPLDSAP 176
Qy 26 ----- 25
Db 177 ARPGHAOPHYVYKROAPERLAORGDSSAPSTCSVPRAGVSTGCALCAAATAARRL 236
Qy 26 -----TERVYLTNLTNGAELLRDPISLGAQPRVHLVK 56
Db 237 HQRSVSKKQVETLVWADAKWVEHQPVQSVYVLTMMVAGLFHDPFSIGNPIHITVR 296

QY	57	MYLITEGAPNITANLTSSLLSVCGWSOTINPEDDTDPGHADLVL	102
Db	297	LVLLSEDEEDLKITHAONTILKSFCKWQKSNKNGDAPLHHDHTAILLTRKDLCAAMNRP	356
QY	103		102
Db	357	CETLGLSHVAGNCQPHRSCSINEDTGLPLAFTVAHELHGSFGIOHDSGNDCEPVGKRP	416
QY	103	YITRF	107
Db	417	INSPOLLYDAAPLTWRSRCSROYITRFLDRGWLCLDDPPAKDIIIDFPPVPPGVLYDVSHQ	476
QY	108		107
Db	477	CRLOYGAYSACFEDMDNVCHTLWCSVGTTCCHSKLDAVDGTRCGENKWCLSGECVPVGR	536
QY	108		107
Db	537	PEAVDGSWSGSAWSICSRSCGWSAERQCTQTPKYKGRYCVGERKRFRLCNLQACP	596
QY	108		107
Db	597	AGRPSPRHVQSHFDAMLYKGOLHTWVPVNDVNPCELHCRPANEFYAKKLRDAVVDGTP	656
QY	108	DLE	110
Db	657	CYQVRASRDLCINGICKNVGCFEIDSGAMEDRCGVCHGNGSTCHTVSTFXRRPRVXY	716
QY	111	LPDGNRQVR	119
Db	717	VDVGLIPAGAREIRIOEAAEANFLARSEDPEKYFLNGWTTQMGDQYVAGTFTYAR	776
QY	120	GYTQ	123
Db	777	RGNWNLSPGPTKEPWIQVPASRPGGSGRGVPRPSTLHGRSPRGVSGSVTEPS	836
QY	124	LGG	126
Db	837	EPGPPAAASTVSPSLKPNLVAAVHRGGQAPLGLGGWRRHLVLMGRPPLTQLLFOES	896
QY	127		126
Db	897	NPGVHYEYTHREAGGHDEVPVPVFSWHPMTKCTVTCRGVQRQNVYCLERQAGPVDE	956
QY	127	ACSP	130
Db	957	EHCPLGRPDDQQRKCEQPCPARWAGEMQLCSCGPGGLSRRRAVLCIRSGLDEQSA	1016
QY	131	SCLTEDTGF	142
Db	1017	LEPPACEHLPRPTETPCNRHVPCPATWAGNWSQCSVTGCGTQRNVLCNTDGTGPCD	1076
QY	143		142
Db	1077	EAQPASEVTCSLPLCRWPLTGLPEGSGSGSSHELNEADFIPLHLAPRPSPASSPKP	1136
QY	143		142
Db	1137	GTWGNATIEEAPELDLPGPVFVDFFYYDYNFINFHEDLSYGPSEEDPLDLAGTDRTPPP	1196
QY	143		142
Db	1197	HSHPAAPSTGSPVPATEPPAAKEEVLGPNWSPWPSQAGSRPPPEQTPGNPLINFLP	1256
QY	143	DLGV	146
Db	1257	EEDTPICAPDLGL	1269

RESULT 13
AAB72283
ID AAB72283 standard; Protein; 997 AA.
XX

AAAB72283;	
14-MAY-2001 (first entry)	
Human ADAMTS-7 amino acid sequence.	
ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif; tumour cachexia; inflammation; dermatosparaxis; EDS-VIIC; angiogenesis; Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human; metastasis; embryogenesis; egg implantation; chromosome 15; ADAMTS-7.	
Homo sapiens.	
WO200111074-A2.	
15-FEB-2001.	
03-AUG-2000; 2000WO-US21223.	
06-AUG-1999; 99US-0369364.	
(CLEV-) CLEVELAND CLINIC FOUND.	
(APTE/) APTE S S.	
(HURS/) HURSKAINEN T L.	
(HIRO/) HIROHATA S.	
Apte SS, Hurskainen TL, Hirohata S;	
WPI; 2001-159978/16.	
N-PSDB; AAF63440.	
Murine and human 'A Disintegrin-like And Metalloprotease domain with Thrombospondin type I motifs' proteins and the nucleic acids encoding them, useful for treating e.g. tumours, inflammation and arthritis .	
Claim 15; Fig 4; 18pp; English.	
This invention relates to murine and human ADAMTS-N (A disintegrin-like and metalloprotease domain with thrombospondin type I motifs) proteins, designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the invention are cDNA sequences encoding the proteins, and antibodies specific for the proteins. The nucleic acid sequences and proteins may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate ADAMTS-N expression. Disorders that may be treated using the nucleic acids, proteins and antibodies include, for example tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage in arthritic (both inflammatory and non-inflammatory) disease, angiogenesis, tumour growth and metastases, and they may also be used for controlling embryogenesis and implantation of fertilised eggs. The present sequence represents human ADAMTS-7. The human ADAMTS-7 gene is located on chromosome 15.	
Sequence 997 AA;	
Query Match 30.7%; Score 237; DB 22; Length 997;	
Best Local Similarity 18.2%; Pred. NO. 0.01;	
Matches 61; Conservative 24; Mismatches 62; Indels 180; Gaps 6;	
QY 1 AAGGTLHLEL-----LVAVG----- 15	
Db 53 AGGSFLSYELWPRALRXRDVSVRRDAPAFYELQYGRRELRLNLTANQHLLAGFVSETRR 112	
QY 16 -----PD-----VFOHQED----- 25	
Db 113 RGLGRAHTRANTPACHLLGEVDPELEGLAAISACDGLKGVFOLSNEYFIEPLDSAP 172	
QY 26 ----- 25	
Db 173 APRGHAQPHVYKQAPERLAQRGSSAPSTCGVQVYPELESRRERWEORQWRRLRR 232	
QY 26 -----TERYVLTNLNIGALLURDPSLGAQFRVHLV 55	

Db 233 LHORSVSEKWCETLVVADAKMVEYHGQPOVESYVLTIMMVAGLFHDPDSIGNPIHITV 292

Qy 56 KVVILTEPCAPNITANLSSLLSVCGWSQTINPEDDTPGHADLVLYITRDLPLDCGN 115

Db 293 RLVLLEDEBEDLKITHADNTLTKSFCKWOKSINMKGDAPLHHDDTAILTRKDL-CAAMN 351

Qy 116 R-QVRGVLTQGGACPTWSCLITDGTGFDLGTI 148

Db 352 RPECITGLSHVAGMCQPHRSCSINEDTGLPLFTV 386

RESULT 14

AAU72897

ID AAU72897 standard; Protein; 1505 AA.

AC AAU72897;

DT 26-FEB-2002 (first entry)

DE Human metalloprotease partial protein sequence #9.

XX Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;

KW vasotropic; antimigraine; analgesic; endocrine; neurotropic; tranquiliser;

KW hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;

KW anorectic; antiinflammatory; aspartyl protease; cysteine protease;

KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;

KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;

KW immune-related disease; cardiovascular disease; neuronal disease;

KW migraine; sexual dysfunction; mood disorder; attention disorder;

KW cognition disorder; hypotension; hypertension; psychotic disorder;

KW dyskinesia; metabolic disorder; inflammatory disorder.

XX Homo sapiens.

OS Homo sapiens.

XX WO200183782-A2.

PN 08-NOV-2001.

XX 04-MAY-2001; 2001WO-US14431.

XX 04-MAY-2000; 2000US-201879P.

PR (SUGEN-) SUGEN INC.

XX Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepel S;

PI Payne V;

XX WPI; 2002-041502/05.

DR N-PSDB; AAS971180.

XX Novel protease polypeptide useful for screening for substances that may

PT be used to treat, e.g., cancers, immune-related diseases,

PT cardiovascular disease, migraine, pain, psychotic and inflammatory

PT disorders -

XX Claim 28; Figure 2F; 232pp; English.

PS The invention relates to an isolated, enriched, or purified protease

CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to

CC screen for substances (S) that may modulate its activity. Administering

CC S (which modulates protease activity in vitro) may be used to treat a

CC disease or disorder selected from cancers (e.g., of tissues, of blood or

CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,

CC brain, ovarian, bladder or kidney), immune-related diseases and

CC disorders, cardiovascular disease, brain or neuronal-associated diseases

CC (e.g., central or peripheral nervous system diseases, migraine, pain,

CC sexual dysfunction, mood disorders, attention disorders, cognition

CC disorders, hypotension, hypertension, psychotic disorders, neurological

CC disorders and dyskinesias), metabolic disorders and inflammatory

CC disorders. (I) may also be useful as a diagnostic tool for a disease or

CC disorder such as those above. AAU72876-AAU72910 represent human

CC protease amino acid sequences of the invention.

SQ Sequence 1505 AA;

Query Match 30.7%; Score 237; DB 23; Length 1505;

Best Local Similarity 5.9%; Pred. No. 0.029;

Matches 57; Conservative 30; Mismatches 51; Indels 830; Gaps 8;

Qy 3 GGIL----- 6

Db 149 GGLTGTGKGNGEYFLEIPMKADGNEYEDGHKPHLYIRODLNNSFLQTLKYCVSSESQI 208

Qy 7 -----HLELLVAVG 15

Db 209 KETSLPFHTYSNMEDLNVMKERVLTGHTSKNVLKDERHRKXKRLISYPRVIEIMVTAD 268

Qy 16 PDVFAQHOEDTERVYVLTNLNIGAEALLRDPISLGAQPRVHLVKMVLITEPEGAPNITANLTS 75

Db 269 AKVSAHGSNLQNYLTILMSIVATYIKDPSIGNLTHIVVVKLVMIHREEGPVINFDGAT 328

Qy 76 SLLSVCGWSQTINPEDDTPGHADLVLYITRDL----- 109

Db 329 TLKNFCSWQOTQNDLDDVHPSHHDTAVLITREDICSSKEKCNMLGLSYLTICDPLQSCF 388

Qy 110 ----- 109

Db 389 INEEKGLISAFTHAELGHTLVQHDDNPRCKEMKVTHVMAPALSPHSPWSWNSCSR 448

Qy 110 -----ELP-----DGNRQ----- 117

Db 449 KYVTBFLDTGYCECLLDKPDDEIYNLPSELPSRVDGNKQCELAFGPGSQMCPHIENICM 508

Qy 118 ----- 117

Db 509 HLWCTSTEKLHKGCTQHVPPADGTDGCGMHCRHGLCVNKETETRPVNGEWGPWEYSS 568

Qy 118 ----- 117

Db 569 CSRTCGGIESATRCNRPEPRNGNYCVGRMRKFRSCNTDSCPKGTODFREKQCSDFNG 628

Qy 118 ----- 117

Db 629 KHLDISGIPSNVRLPRYSGIGTKDKRCLYCOVAGTNYFYLLKDMVEDGTGCTETHDIC 688

Qy 118 ----- 117

Db 689 VQGQCMAGCDHVLNSSAKIDKCGVCGDGNSSCKTITGVFNSSHYGVNVVVKIPAGATNV 748

Qy 118 -----VRGVTLQGA----- 127

Db 749 DIRQVSYSCQPDSDSYLAUSDAGNLFNGNFFLLSTSKKEINVQGTTRTVIEYSGSNNAVER 808

Qy 128 ----- 127

Db 809 INSTNRQEKELILQVLCVGNLYNPDVHYSFNIPLEERSDMFTWDPYGPWEGCTKMCQGLQ 868

Qy 128 ----- 127

Db 869 RRNITCIHKSDHSVSDKEDHLPLPSFVTQSCNTDCELRMHVIGKSECSOCCGYRTL 928

Qy 128 -----CS----- 129

Db 929 DIHKMYSIHGOTVQDDHDCDQLKPTQBELCHGNCVFTRWHHYSEWSQCSRSCCGGER 988

Qy 130 -----PTWS-----CLIT----- 137

Db 989 SRESYCMNFGHRLADNECQELSRVTRENCNPFSCPSWAASESECLVTCGKTQKQVW 1048

Qy 138 -----EDT 140

Db 1049 CQLNVDLSDGFCNSSTKPSLSPCELHPCASQWQVGPWGPCTTTTCGHGYQMRDVKCVNEL 1108

Qy 138 -----EDT 140

Db 1109 ASAVLEDT 1116

Job time : 70 secs

```

RESULT 15
AAB74944
ID AAB74944 standard; Protein; 1686 AA.
XX
AC AAB74944;
XX
DT 02-JUL-2001 (first entry)
XX
DE Human ADAM type metal protease MDTs1 protein SEQ ID NO:1.
XX
KW Human; a disintegrin and metalloprotease type metal protease; MDTs1;
KW MDTs2; MDTs3; ADAM type metal protease; cytosolic; antiarthritic;
KW cancer; arthritis; arthrosis; deformans.
XX
OS Homo sapiens.
XX
PN JP2001008687-A.
XX
PD 16-JAN-2001.
XX
PF 25-JUN-1999; 99JP-0180973.
XX
PR 25-JUN-1999; 99JP-0180973.
XX
PA (YAMA) YAMANOUCHI PHARM CO LTD.
XX
DR WPI; 2001-285362/30.
DR N-PSDB; AAF82149.
XX
PT New metal protease and metal protease gene, for use as a drug for
PT cancers, arthritis and arthrosis deformans.
XX
PS Claim 1; Page 12-17; 31pp; Japanese.
XX
CC The present sequence represents a disintegrin and metalloprotease (ADAM)
CC type metal protease designated MDTs1, isolated from human. MDTs proteins
CC have cytosolic and antiarthritic activities. They can be used as a drug
CC for cancers, arthritis and arthrosis deformans.
XX
SQ Sequence 1686 AA;

Query Match 30.7%; Score 237; DB 22; Length 1686;
Best Local Similarity 18.2%; Pred. No. 0.038;
Matches 61; Conservative 24; Mismatches 62; Indels 198; Gaps 6;

QY 1 AAGGILHLEL-----LVAVG----- 15
DB 53 AGGSFSLYELWPRALRKRDVSVRDAPAFYELQYRGRELNFNLTAHQHLLAPGVSETRR 112
QY 16 -----PD-----VFOHQED----- 25
DB 113 RGLGRAHRAHTPACHLLGEVDPEGLAAISACDGLKGVFQLSNEDYFIEPLDSAP 172
QY 26 ----- 25
DB 173 ARPQHAQPHVYVYKQAPERLAQRGDSAPSTCGVQVYPELESRRERWEQOQWRRRLRR 232
QY 26 -----TERYVLTNLNIGAEILLRDPISLGAQFRVHLV 55
DB 233 LHQSVSKKQVETLVADAKMVEYHGQVQVESYVLTIMNVAGLFHDPISGNPIHITIV 292
QY 56 KMWILTEPEGAPNITANITSSLLSCGWSQTNPEDDTPGHADLVLYITRFDLELPDGN 115
DB 293 RLVLLDEEEDLKITHHADNTLKSFCWKQKSNMKGDAPLHHDFAILLTRKDL-CAAMN 351
QY 116 R--QVRGVTQLGACSPWCLITEDTGTDLGVTI 148
DB 352 RPECETGLGSHVAGMCMQPHRSCSINEDTGLPLAFTV 386

```

Search completed: March 20, 2003, 12:37:50

GenCore version 5.1.4 ps_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:31:32 ; Search time 49.6216 Seconds
(without alignments)
564.722 Million cell updates/sec

Title: SEQID_15
Perfect score: 716
Sequence: 1 AVGPVFOAQHEDTERVLT.....SPTWSCLITGFDLGVIT 136

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_proteob.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	716	100.0	1427	4 Q96L37	Q96L37 homo sapien
2	201	28.1	2165	5 Q19791	Q19791 caenorhabdi
3	194	27.1	1054	5 Q9W493	Q9W493 drosophila
4	173.5	24.2	950	4 Q8TE58	Q8TE58 homo sapien
5	160	22.3	1159	4 Q8TEY8	Q8TEY8 homo sapien
6	154.5	21.6	1223	4 Q8WX58	Q8WX58 homo sapien
7	154.5	21.6	1223	4 Q8TE55	Q8TE55 homo sapien
8	153.5	21.4	1095	4 Q8TE56	Q8TE56 homo sapien
9	147.5	20.6	1072	4 Q8TE57	Q8TE57 homo sapien
10	138.5	19.3	1207	4 Q8TE59	Q8TE59 homo sapien
11	138	19.3	1091	5 Q9W126	Q9W126 drosophila
12	137	19.1	790	5 Q8T458	Q8T458 drosophila
13	128.5	17.9	1081	4 Q8TE60	Q8TE60 homo sapien
14	128.5	17.9	1229	5 Q9VF61	Q9VF61 drosophila
15	128.5	17.9	1688	5 Q8SX60	Q8SX60 drosophila
16	112.5	15.7	872	5 Q22580	Q22580 caenorhabdi

17	85	11.9	1062	5 Q19204	Q19204 caenorhabdi
18	81.5	11.4	541	16 Q9F3J1	Q9F3J1 streptomyc
19	76	10.6	802	11 Q91Y03	Q91Y03 mus musculu
20	75.5	10.5	152	10 Q04205	Q04205 arabidopels
21	75.5	10.5	509	5 Q9NDL4	Q9NDL4 caenorhabdi
22	75.5	10.5	509	5 Q20930	Q20930 caenorhabdi
23	74	10.3	755	5 Q9VF30	Q9VF30 drosophila
24	73.5	10.3	1444	5 Q17591	Q17591 caenorhabdi
25	72.5	10.1	355	4 Q9H6J3	Q9H6J3 homo sapien
26	72	10.1	531	4 Q9UH32	Q9UH32 homo sapien
27	72	10.1	1190	5 Q9V7E5	Q9V7E5 drosophila
28	71.5	10.0	518	10 Q9LJ22	Q9LJ22 oryza sativ
29	71.5	10.0	723	4 Q9UF23	Q9UF23 homo sapien
30	71.5	10.0	963	4 Q9F202	Q9F202 homo sapien
31	71	9.9	236	16 Q9KNV5	Q9KNV5 vibrio chol
32	71	9.9	254	8 Q94X42	Q94X42 bemisia tab
33	71	9.9	443	16 Q9HT65	Q9HT65 pseudomonas
34	71	9.9	726	4 Q9Y416	Q9Y416 homo sapien
35	71	9.9	906	4 Q8WVS2	Q8WVS2 homo sapien
36	71	9.9	1104	4 Q9BSJ8	Q9BSJ8 homo sapien
37	70.5	9.8	227	2 Q9L8E8	Q9L8E8 vibrio harv
38	70.5	9.8	266	17 Q57843	Q57843 pyrococcus
39	70.5	9.8	275	5 Q9VP95	Q9VP95 drosophila
40	70.5	9.8	388	16 Q9RK22	Q9RK22 streptomyc
41	70.5	9.8	539	3 P78832	P78832 echizoaecch
42	70.5	9.8	593	11 Q9Z5M0	Q9Z5M0 mus musculu
43	70.5	9.8	785	16 Q98KF6	Q98KF6 rhizobium l
44	70.5	9.8	793	11 Q91X21	Q91X21 mus musculu
45	70	9.8	358	2 Q9X5K8	Q9X5K8 streptomyc

ALIGNMENTS

RESULT 1

Q96L37	PRELIMINARY;	PRT; 1427 AA.
ID	Q96L37	
AC	Q96L37;	
DT	01-DEC-2001 (TREMBLrel. 19, Created)	
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE	Von Willebrand factor-cleaving protease precursor.	
GN	ADAMTS13.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=LIVER;	
RX	PubMed=11557746;	
RA	Zheng X., Chung D., Takayama T.K., Majerus E.M., Sadler J.E.,	
RA	Fujikawa K.;	
RT	Structure of von Willebrand Factor-cleaving Protease (ADAMTS13), a	
RT	Metalloprotease Involved in Thrombotic Thrombocytopenic Purpura.;	
RL	J. Biol. Chem. 276:41059-41063(2001).	
DR	EMBL; AY055376; AAL17652.1;	
DR	MEROPS; M12.241;	
DR	InterPro; IPR001590; Reprolysin.	
DR	InterPro; IPR000884; TSPI.	
DR	InterPro; IPR000130; Zn_Mtpeptdase.	
DR	Pfam; PF01421; Reprolysin; 1.	
DR	Pfam; PF00090; tsg_1; 4.	
DR	PROSITE; PS0215; ADAM_MEPRO; 1.	
DR	PROSITE; PS0092; TSPI; 1.	
DR	PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.	
DR	Protease; Signal.	
KW	SIGNAL	1 33
FT	CHAIN	75 1427
FT	VON WILLEBRAND FACTOR-CLEAVING PROTEASE.	
SQ	SEQUENCE	1427 AA; 153632 MW; EB1BC3AABCI4442 CRC64;

Query Match 100.0%; Score 716; DB 4; Length 1427;
Best Local Similarity 100.0%; Pred. No. 2.6e-69;


```

DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reptolysin; 1.
DR SMART; PF00090; tsp_1; 2.
DR SMART; SM00209; TSP1; 2.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSP1; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 1054 AA; 118616 MW; DC15455555CB6212 CRC64;

Query Match
Best Local Similarity 27.1%; Score 194; DB 5; Length 1054;
Matches 44; Conservative 21; Mismatches 66; Indels 4; Gaps 2;

QY 2 VGPDVFAHQEDTERYVLTNLNIGALLRDPGLGAQFRVHLVVKMVLTEPEG--APNITA 59
DB 332 VADATMSAFHRDLNGYLLTINMWSALYKDPISGNSIEIVVRIQLDEESQLQLNLTQ 391
QY 60 NLTSSLLSCGWSQTINPEDDTPGHADLVLYITRFDLELPGNRQVRGVTLGGACSP 119
DB 392 NAQKVLDRFCSQWHLKNGSEKDPHHHDVAIIITR--KNICANNMTGLANVGCMCKPK 449
QY 120 WSLCLITDGTGDLGVTI 136
DB 450 QSCSVNEDNGIMLSHTI 466

RESULT 4
Q8TE58 PRELIMINARY; PRT; 950 AA.
AC Q8TE58;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Metalloprotease disintegrin 15 with thrombospondin domains.
DE ADAMTS15.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.;
RL Gene 283.49-62(2002).
DR EMBL; AJ315733; CAC86014.1; -.
KW Integrin; Protease.
SQ SEQUENCE 950 AA; 103286 MW; 5DFEE18285CCCC3B CRC64;

Query Match
Best Local Similarity 24.2%; Score 173.5; DB 4; Length 950;
Matches 39; Conservative 19; Mismatches 67; Indels 1; Gaps 1;

QY 10 HQEDTERYVLTNLNIGALLRDPGLGAQFRVHLVVKMVLTEPEGAPNITANLTSSLLSY 69
DB 234 HGADLEHYLLTLTATARLYRHPHSILPINIVVVKLLLRDRSDGPKVTGNAALTLRNFC 293
QY 70 GWSQTINPEDDTPGHADLVLYITRFDLELPGNRQVRGVTLGGACSPWSCLITDGTG 129
DB 294 AWOKLKNVSKHPEYWDITAILFTRODL-CGATTCDTLGMADVGTWCDPKRSCSVIEDDG 352
QY 130 FDLGVT 135
DB 353 LPSAFT 358

RESULT 5
Q8TEY8 PRELIMINARY; PRT; 1159 AA.
AC Q8TEY8;

```

```

DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE ADAMTS14.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21139041; PubMed=11741898;
RA Collige A., Vandenbergh I., Thiry M., Lambert C.A., Van Beeumen J.,
RA Li S.W., Prockop D.J., Lapiere C.M., Nussgens B.V.;
RT "Cloning and Characterization of ADAMTS-14, a Novel ADAMTS Displaying
RT High Homology with ADAMTS-2 and ADAMTS-3.";
RL J. Biol. Chem. 277:5756-5766(2002).
DR EMBL; AF366351; AAL79814.1; -.
SQ SEQUENCE 1159 AA; 127336 MW; ASB130149BF7FF34 CRC64;

Query Match
Best Local Similarity 22.3%; Score 160; DB 4; Length 1159;
Matches 41; Conservative 25; Mismatches 65; Indels 6; Gaps 3;

QY 2 VGPDVFAHQEDTERYVLTNLNIGALLRDPGLGAQFRVHLVVKMVLTEPEGAPNI-TA 59
DB 200 VDDSVVRFHGKHVQNYVLTLMNIVDEIYHDESLGWHINIALVRLIMVGYRSLSLIERG 259
QY 60 NLTSSLLSCGWSQTINPEDDTPGHADLVLYITRFDLELPGNRQVRGVTLGGACSP 119
DB 260 NPSRSLEQVCRWAHSQORQDPSHAHHHVVLTTRQDF----GPSGMQGYAPVTGWHCHPL 315
QY 120 WSLCLITDGTGDLGVTI 136
DB 316 RSCALNHEDGFSFAFVI 332

RESULT 6
Q8WX58 PRELIMINARY; PRT; 1223 AA.
AC Q8WX58;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE A disintegrin-like and metalloprotease with thrombospondin type 1
DE motif 14 precursor.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638061; PubMed=11796938;
RA Bolz H., Ramirez A., von Brederlow B., Kubisch C.;
RT "Characterization of ADAMTS14, a novel member of the ADAMTS
RT metalloproteinase family.";
RL Biochim. Biophys. Acta 1522:221-225(2001).
DR EMBL; AF358666; AAL40229.1; -.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reptolysin.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF00090; TSP1; 4.
DR SMART; SM00209; TSP1; 4.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSP1; 1.
DR Signal; Integrin; Protease; Metalloprotease.
FT SIGNAL 1 22 POTENTIAL.
SQ SEQUENCE 1223 AA; 133871 MW; 36394AC4D92F170F CRC64;

Query Match
Best Local Similarity 21.6%; Score 154.5; DB 4; Length 1223;

```

```

Best Local Similarity 30.7%; Pred. No. 7.2e-08;
Matches 42; Conservative 23; Mismatches 63; Indels 9; Gaps 4;

QY 2 VGPDVFOAH-QEDTERVYLTNINIGAEILLRDPISLGAQFRVHLVKMVLTEPEGAPNI-TA 59
DB 267 VDDSVVRFHKGHEVQNYVLTLMNIVDEIYHDESGLGHINIALVRLIMVGYRQSLIERG 326
QY 60 NLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGDGNRQVRGVTLGGACSP 119
DB 327 NPSRSLEQVCRWAHSQQRDPHAEHHDHVVELTRQDFG-PSGYAPVTGM-----CHPL 379

QY 120 WSLCLITDTGFDLGVTI 136
DB 380 RSCALNHEDGFSSAFVI 396

RESULT 7
QY 8TE55 PRELIMINARY; PRT; 1223 AA.
AC 8TE55;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Metalloprotease-disintegrin protease.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 11_
RP SEQUENCE FROM N.A.
RC TISSUE=FETAL LUNG;
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ315734; CAC86015.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1223 AA; 133925 MW; D585B6593977ED15 CRC64;

Query Match 21.6%; Score 154.5; DB 4; Length 1223;
Best Local Similarity 30.7%; Pred. No. 7.2e-08;
Matches 42; Conservative 23; Mismatches 63; Indels 9; Gaps 4;

QY 2 VGPDVFOAH-QEDTERVYLTNINIGAEILLRDPISLGAQFRVHLVKMVLTEPEGAPNI-TA 59
DB 267 VDDSVVRFHKGHEVQNYVLTLMNIVDEIYHDESGLGHINIALVRLIMVGYRQSLIERG 326
QY 60 NLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGDGNRQVRGVTLGGACSP 119
DB 327 NPSRSLEQVCRWAHSQQRDPHAEHHDHVVELTRQDFG-PSGYAPVTGM-----CHPL 379

QY 120 WSLCLITDTGFDLGVTI 136
DB 380 RSCALNHEDGFSSAFVI 396

RESULT 8
QY 8TE56 PRELIMINARY; PRT; 1095 AA.
AC 8TE56;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Metalloprotease disintegrin 17, with thrombospondin domains.
GN ADAMTS17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```

```

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ315735; CAC86016.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1095 AA; 121099 MW; ASC6540484754D5D CRC64;

Query Match 21.4%; Score 153.5; DB 4; Length 1095;
Best Local Similarity 28.3%; Pred. No. 8e-08;
Matches 41; Conservative 24; Mismatches 67; Indels 13; Gaps 3;

QY 5 DVFOAH-QEDTERVYLTNINIGAEILLRDPISLGAQFRVHLVKMVLTEPEGAPNITANLTS 63
DB 243 DMVQYHGAEAAQRFILTMNVMNVMFQHSGLGINKINTQVTLVLLRORPAKLSIGHGER 302
QY 64 SLLSVCGW-----SQTINPEDDTPGHADLVLYITRFDLEL-PDGNRQVRGV 111
DB 303 SLESFCHWQNEEYCGARYLGNQVPGKDDPPLVDAAVFVTRTDFCVHKDEPCDTVGAIY 362
QY 112 LGGACSPWTSCLITDTGFDLGVTI 136
DB 363 LGGVCSAKKCKVLAEDNGLNLAFTI 387

RESULT 9
QY 8TE57 PRELIMINARY; PRT; 1072 AA.
AC 8TE57;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Metalloprotease disintegrin 16 with thrombospondin type I motif.
GN ADAMTS16.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 11_
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ315734; CAC86015.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1072 AA; 119491 MW; 081EEFF78F47D061 CRC64;

Query Match 20.6%; Score 147.5; DB 4; Length 1072;
Best Local Similarity 31.2%; Pred. No. 3.5e-07;
Matches 43; Conservative 22; Mismatches 66; Indels 7; Gaps 3;

QY 2 VGPDVFOAH-QEDTERVYLTNINIGAEILLRDPISLGAQFRVHLVKMVLTEPEGAPNITAN 60
DB 298 VDKKMQNHGHENITTYVLTILNVMVSALFKDGTIGGNINIAIVGLILDEOPGLVISHH 357
QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLV--LYITRFDLELPGDGNRQVRGVTLGGACSP 118
DB 358 ADHTLSSFCQWSQSLGKQDGRHDAILLTGLDICSWKNEPCD-----TLGFAPISGMSCK 413
QY 119 TWSCLITDTGFDLGVTI 136
DB 414 YRSCTINEDTGLGLAFTI 431
```



```

RESULT 10
Q8T559 PRELIMINARY; PRT; 1207 AA.
AC Q8T559;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE ADAMTS-19.
GN ADAMTS19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT diintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ311904; CAC84565.1; --
SQ SEQUENCE 1207 AA; 134061 MW; AF36F6BF586GFDE2 CRC64;

Query Match 19.3%; Score 138.5; DB 4; Length 1207;
Best Local Similarity 30.1%; Pred. No. 4e-06;
Matches 44; Conservative 19; Mismatches 70; Indels 13; Gaps 4;

QY 4 PDVFOAHQED-TRYVLTNLNIGAEILLRDPGLGQAFRVHLVKNVILTEPEGAPNITANLT 62
Db 335 PAMVSYHGADAARRFILTILNMVFNLFQKSLGVQVNLRIKILLIHETPELYIGHGE 394

QY 63 SLLSVCGV-----SQTINPEDDTPGH-----ADLVLYITRFDLLE-PDGNRQVRGVT 110
Db 395 KMLESPCKQHEBFEGKKNKHLEMINWGEDMTSVDAAILITRKDFCVHKDEPDCVTGIA 454

QY 111 QLGGACSPWTSCLITDGTGDLGVTI 136
Db 455 YLSGMCSEKRCIIAEDNGNLNLAFTI 480

RESULT 11
Q9W126 PRELIMINARY; PRT; 1091 AA.
AC Q9W126;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE CG3622 protein.
GN CG3622.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherter S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Branton R.C., Rogers Y.-H.C., Blazej R.G., Champ M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.H., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Shandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Broksstein P., Brotter P.,
RA Butris K.C., Bueam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

```

```

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Foaler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Morkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003459; AAF46905.1; --
DR FlyBase; FBGN0034778; CG3622.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000130; Zn_MTPetpdae.
DR Pfam; PF01421; Reprolysin; 1.
DR PROSITE; PS00215; ADAM_MEPHO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 1091 AA; 124074 MW; 5B9673B2F12565BE CRC64;

Query Match 19.3%; Score 138; DB 5; Length 1091;
Best Local Similarity 31.0%; Pred. No. 4e-06; 57; Indels 18; Gaps 8;
Matches 45; Conservative 25; Mismatches 25;

QY 2 VGPDVFOAHQED-TRYVLTNLNIGAEILLRDPGLGQAFRVHLVKNVILTEPEGAPNITANLT 49
Db 294 VPSDLY-AHWKQNFPTNTESKVSFLAMIN-GVQLLYHHPTLGRINPVLKRLKWSW 351

QY 50 EPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLLEP---GNRQV 106
Db 352 DPPGLVR-SRDVENYLSFKQKQLNPFSDADPLHYDHALVLTGLDLYTDGKANSQV 410

QY 107 RGVTLGGACSPWTSCLITDGTGFD 131
Db 411 VGMATVKGMCTSIYSCTINEAKHFE 435

RESULT 12
Q8T458 PRELIMINARY; PRT; 790 AA.
AC Q8T458;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE AT15733p.
GN CG3622.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Drenek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungali C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:31:05 ; Search time 70.7568 Seconds
(without alignments)
256.118 Million cell updates/sec

Title: SEQID_15

Perfect score: 716

Sequence: 1 AVGPDPVQAHQEDTERVLT.....SPTWSCLITDGTGDLGVTI 136

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	716	100.0	148	AAE24448	Human Von Willebra
2	716	100.0	242	AAU97641	Human aggrecanase
3	716	100.0	933	AAU79217	Human ADAM-TS-like
4	716	100.0	1353	AAE24449	Human Von Willebra
5	716	100.0	1416	AB04153	Human ADAMTS-M pol
6	716	100.0	1427	AAE24450	Human Von Willebra
7	713	99.6	203	AAE3826	Amino acid sequenc
8	713	99.6	1120	AAE63829	Amino acid sequenc
9	529	73.9	118	AAE42668	Human ORFX ORF2432
10	212.5	29.7	997	AAE72283	Human ADAMTS-7 ami

11	212.5	29.7	1686	22	AAE74944	Human ADAM type me
12	212.5	29.7	1686	22	AAE00913	Human 27875 ADAM-T
13	212.5	29.7	1686	22	AAE00934	Human 27875 ADAM-T
14	212.5	29.7	1784	21	AAE41379	Human ORFX ORF1143
15	210	29.3	1505	23	AAU72897	Human metalloprote
16	206.5	28.8	1690	22	AAE86949	Human metalloprote
17	201	28.1	2150	21	AAE53898	Amino acid sequenc
18	201	28.1	2165	22	AAE90617	Human secreted pro
19	194	27.1	1054	22	ABE0410	Drosophila melanog
20	189.5	26.5	1073	21	AAE21264	Human metalloprote
21	187.5	26.2	947	22	AAE86950	Human metalloprote
22	187.5	26.2	1602	23	ABG30702	Human aggrecanase
23	187.5	26.2	1629	23	ABG30703	Human aggrecanase
24	187.5	26.2	1629	23	AAO14448	Human ADAMTS-Si pr
25	187.5	26.2	1916	23	AAE19173	Human protease, PR
26	187.5	26.2	1935	23	AAU72896	Human metalloprote
27	184	25.7	1882	22	AAE72286	Human ADAMTS-9 ami
28	184	25.7	1934	22	AAE72301	Human ADAMTS-9 alt
29	178.5	24.9	727	20	AAW78435	Human ADAMTS-1 pro
30	178.5	24.9	950	20	AAE49501	Human METH1 protei
31	178.5	24.9	950	20	AAE73549	Human ADAM-type me
32	178.5	24.9	950	22	AAE50002	Human METH1. Homo
33	178.5	24.9	967	19	AAE80285	Human integrin lig
34	178.5	24.9	967	19	AAE04142	Human Tango-71 pro
35	178.5	24.9	968	22	AAE50011	Mouse metalloprote
36	177.5	24.8	896	21	AAE21265	Human protease-rel
37	176	24.6	317	22	AAE03579	Human protease-rel
38	176	24.6	356	22	AAE03581	Human protease-rel
39	176	24.6	438	22	AAE03582	Human protease-rel
40	176	24.6	468	22	AAE03574	Human protease-rel
41	176	24.6	507	22	AAE03576	Human protease-rel
42	176	24.6	589	22	AAE03577	Human protease-rel
43	176	24.6	757	22	AAE03583	Human protease-rel
44	176	24.6	859	22	AAE72282	Human ADAMTS-6 ami
45	176	24.6	908	22	AAE03572	Human protease-rel

ALIGNMENTS

RESULT 1
AAE24448
ID AAE24448 standard; Protein; 148 AA.
XX
AC AAE24448;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human Von Willebrand factor-cleaving protease fragment #1.
XX
KW Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;
KW transgenic animal; immunisation; thromboembolic disease; preclampsia;
KW thrombotic thrombocytopenic purpura; TTP; Henoch-Schonlein purpura;
KW thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
KW transgenic; anticoagulant.
OS Homo sapiens.
XX
XX WO200242441-A2.
XX
PD 30-MAY-2002.
XX
XX 20-NOV-2001; 2001WO-EP13391.
XX
XX 22-NOV-2000; 2000US-0721254;
XX 12-APR-2001; 2001US-083328;
XX (BAXT) BAXTER AG.
XX
XX Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
XX Scheifflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
XX Zimmermann K, Voelkel D;
XX

DR WPI; 2002-479950/51.
 DR N-PSDB; AAD39331.
 PT Novel isolated or substantially purified Von Willebrand factor-cleaving
 PT protease, useful for producing preparation for therapy of thrombosis
 PT and thromboembolic disease such as thrombotic thrombocytic purpura -
 XX
 PS Claim 1; Fig 3; 93pp; English.
 CC
 CC The invention relates to an isolated or substantially pure Von Willebrand
 CC factor-cleaving protease (VWF-cp) polypeptide. VWF-cp is useful for
 CC purifying VWF which involves providing VWF-cp as a ligand, contacting a
 CC solution comprising VWF with the polypeptide ligand under conditions
 CC where VWF is bound to the ligand and recovering from the ligand purified
 CC VWF. VWF-cp is useful for producing anti-VWF cp polypeptide antibodies
 CC which involves immunising an animal with VWF-cp and isolating the anti-
 CC VWF cp polypeptide antibodies from the animal. VWF-cp is useful for
 CC producing a preparation of prophylaxis and therapy of thrombosis and
 CC thromboembolic disease such as thrombotic thrombocytic purpura (TTP),
 CC Henoch-Schönlein purpura, preeclampsia, neonatal thrombocytopenia or
 CC haemolytic-uraemic syndrome. VWF-cp can also be used for processing
 CC plasmatic or recombinantly produced VWF. The invention is useful for
 CC construction expression systems and generating transgenic animals which
 CC express the polypeptide in vivo. The present sequence is human VWF-cp
 CC fragment.
 CC
 CC Sequence 148 AA;
 SQ
 Query Match 100.0%; Score 716; DB 23; Length 148;
 Best Local Similarity 100.0%; Pred. No. 3e-78;
 Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AVGPVFOAHOEDTERVYLTNLNIGAEELRDPGLGQFRVHLVKNVILTEPEGAPNITAN 60
 DB 13 AVGPVFOAHOEDTERVYLTNLNIGAEELRDPGLGQFRVHLVKNVILTEPEGAPNITAN 72
 QY 61 LTSSLLSVCGWSQITINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTOLGGACSPW 120
 DB 73 LTSSLLSVCGWSQITINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTOLGGACSPW 132
 QY 121 SCLITEDTGFDLGVTI 136
 DB 133 SCLITEDTGFDLGVTI 148
 RESULT 2
 AAU97641
 ID AAU97641 standard; Protein; 242 AA.
 AC AAU97641;
 XX
 XX 27-AUG-2002 (first entry)
 DT
 DE Human aggrecanase protein.
 XX
 XX Aggrecanase; osteopathic; antiarthritic; antiinflammatory; human;
 KW cartilage; osteoarthritis; inflammatory disease; enzyme.
 KW
 XX Homo sapiens.
 OS
 XX WO200233093-A2.
 PN
 XX 25-APR-2002.
 PD
 XX 17-OCT-2001; 2001WO-032458.
 PF
 XX 18-OCT-2000; 2000US-241469P.
 PR
 XX (GEMY) GENETICS INST INC.
 PA
 XX Racie LA, Twine NC, Agostino MJ, Wolfman NM, Morris EA;
 PI
 XX WPI; 2002-452389/48.
 XX
 DR

DR N-PSDB; ABK52579.
 XX
 XX Novel purified aggrecanase polypeptide useful for developing inhibitors
 PT and antibodies to the aggrecanase polypeptide, which are useful for
 PT treating aggrecanase-associated condition such as osteoarthritis -
 XX
 PS Claim 1; Page 33-34; 41pp; English.
 CC
 CC This invention relates to the cDNA and protein sequences of a novel
 CC human aggrecanase polypeptide. The protein of the invention may be
 CC used to inhibit the proteolytic activity of aggrecanase, or to inhibit
 CC the aggrecanase-mediated cleavage of aggrecan in cartilage. The
 CC protein of the invention is useful for developing inhibitors of
 CC aggrecanase protein. The cDNA sequence encoding the aggrecanase protein
 CC of the invention is useful for designing probes for obtaining DNA
 CC sequences encoding other aggrecanase molecules. The cDNA sequence is
 CC also useful for detecting mRNA encoding aggrecanase in a given cell
 CC population, and thus for detecting or diagnosing genetic disorders
 CC involving the aggrecanase, or disorders involving cellular, organ or
 CC tissue disorders in which aggrecanase is irregularly transcribed or
 CC expressed. The DNA sequences may also be useful for preparing vectors
 CC for gene therapy applications. An inhibitor of the protein is useful
 CC in treating conditions characterised by degradation of articular
 CC cartilage, by blocking the enzyme's proteolytic activity. An
 CC aggrecanase protein inhibitor and a method for inhibition of its activity
 CC are useful for treating various aggrecanase-associated conditions
 CC including osteoarthritis and other inflammatory diseases. The
 CC present sequence represents the human aggrecanase protein of the
 CC invention.
 CC
 CC Sequence 242 AA;
 SQ
 Query Match 100.0%; Score 716; DB 23; Length 242;
 Best Local Similarity 100.0%; Pred. No. 6.2e-78;
 Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AVGPVFOAHOEDTERVYLTNLNIGAEELRDPGLGQFRVHLVKNVILTEPEGAPNITAN 60
 DB 54 AVGPVFOAHOEDTERVYLTNLNIGAEELRDPGLGQFRVHLVKNVILTEPEGAPNITAN 113
 QY 61 LTSSLLSVCGWSQITINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTOLGGACSPW 120
 DB 114 LTSSLLSVCGWSQITINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTOLGGACSPW 173
 QY 121 SCLITEDTGFDLGVTI 136
 DB 174 SCLITEDTGFDLGVTI 189
 RESULT 3
 AAU79217
 ID AAU79217 standard; Protein; 933 AA.
 AC AAU79217;
 XX
 XX 15-JUL-2002 (first entry)
 DT
 DE Human ADAM-TS-like protein.
 XX
 XX Human; ADAM-TS-like protein; cardiovascular disorder; angina;
 KW vascular system; congestive heart failure; myocardial infarction;
 KW ischemic heart disease; arrhythmia; hypertensive vascular disease;
 KW secondary peripheral arterial hypertension; peripheral vascular disease; embolism;
 KW chronic peripheral arterial occlusive disease; acute arterial thrombosis;
 KW inflammatory vascular disorder; chronic obstructive pulmonary disease;
 KW liver disorder.
 XX
 OS Homo sapiens.
 XX WO200226999-A2.
 PN
 XX 04-APR-2002.
 PD
 XX
 XX

```

PF 26-SEP-2001; 2001WO-EP11124.
XX
XX 28-SEP-2000; 2000US-235881P.
PR 25-JUL-2001; 2001US-307393P.
XX
XX (FARB ) BAYER AG.
XX
XX Xiao Y;
XX
XX WPI: 2002-383274/41.
DR N-PSDB; ABK49356.
XX
XX New purified human ADAM-TS-like protein, useful for identifying
PT modulators of protein activity for treating cardiovascular or liver
PT disorder or chronic obstructive pulmonary disease -
XX
XX Claim 25; Fig 2; 106pp; English.
XX
XX The invention relates to a human ADAM-TS-like protein and the
CC polynucleotide encoding it. The protein of the invention is useful for
CC treating cardiovascular disorders including diseases of the heart and
CC vascular system, such as congestive heart failure, myocardial infarction,
CC ischaemic heart diseases (e.g., stable angina, unstable angina), atrial
CC and ventricular arrhythmia, hypertensive vascular diseases (e.g., all
CC kinds of secondary arterial hypertension), and peripheral vascular
CC diseases (e.g., chronic peripheral arterial occlusive disease, acute
CC arterial thrombosis and embolism, inflammatory vascular disorders, etc),
CC liver disorders and chronic obstructive pulmonary disease. The sequences
CC are useful in diagnostic assays for detecting diseases and abnormalities
CC or susceptibility to diseases and abnormalities related to the presence
CC of mutations in the nucleic acid sequences which encode the protein. The
CC sequences are also useful for modulating ADAM-TS-like protein activity in
CC a disease condition. This sequence represents the human ADAM-TS-like
CC protein.
XX
XX Sequence 933 AA;
SQ
Query Match 100.0%; Score 716; DB 23; Length 933;
Best Local Similarity 100.0%; Pred. No. 4.4e-77;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AVGPDVFAHQEDTERVLTNLNIGAEELRDPGLGAQFRVHLVKMVLTEPEGAPNITAN 60
DB 90 AVGPDVFAHQEDTERVLTNLNIGAEELRDPGLGAQFRVHLVKMVLTEPEGAPNITAN 149
OY 61 LTSSLLSVCGWSQTNIPEDDTPGHADLVLYITRFDLELDPGNRQVRGVTLGGACSPW 120
DB 150 LTSSLLSVCGWSQTNIPEDDTPGHADLVLYITRFDLELDPGNRQVRGVTLGGACSPW 209
OY 121 SCLITEDTGFDLGVTI 136
DB 210 SCLITEDTGFDLGVTI 225
RESULT 4
AAE24449
ID AAE24449 standard; Protein; 1353 AA.
XX
XX AAE24449;
XX
XX 04-OCT-2002 (first entry)
XX
XX Human Von Willebrand factor-cleaving protease fragment #2.
XX
XX Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;
XX transgenic animal; immunisation; thromboembolic disease; preclampsia;
XX thrombotic thrombocytopenic purpura; TTP; Henoch-Schonlein purpura;
XX thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
XX transgenic; anticoagulant.
XX
XX Homo sapiens.
XX
XX WO200242441-A2.
PN
30-MAY-2002.
XX
XX 20-NOV-2001; 2001WO-EP13391.
XX
XX 22-NOV-2000; 2000US-0721254.
PR 12-APR-2001; 2001US-083328.
XX
XX (BAXT ) BAXTER AG.
XX
XX Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
PI Scheiflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
XX Zimmermann K, Voelkel D;
XX WPI: 2002-479950/51.
XX
XX Novel isolated or substantially purified Von Willebrand factor-cleaving
PT protease, useful for producing preparation for therapy of thrombosis
PT and thromboembolic disease such as thrombotic thrombocytopenic purpura -
XX
XX Claim 1; Page 64-68; 93pp; English.
XX
XX The invention relates to an isolated or substantially pure Von Willebrand
CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
CC solution comprising vWF with the polypeptide ligand under conditions
CC where vWF is bound to the ligand and recovering from the ligand purified
CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
CC which involves immunising an animal with vWF-cp and isolating the anti-
CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
CC producing a preparation of prophylaxis and therapy of thrombosis and
CC thromboembolic disease such as thrombotic thrombocytopenic purpura (TTP),
CC Henoch-Schonlein purpura, preclampsia, neonatal thrombocytopenia or
CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
CC plasmatic or recombinantly produced vWF. The invention is useful for
CC construction expression systems and generating transgenic animals which
CC express the polypeptide in vivo. The present sequence is human vWF-cp
CC fragment.
XX
XX Sequence 1353 AA;
SQ
Query Match 100.0%; Score 716; DB 23; Length 1353;
Best Local Similarity 100.0%; Pred. No. 7.6e-77;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AVGPDVFAHQEDTERVLTNLNIGAEELRDPGLGAQFRVHLVKMVLTEPEGAPNITAN 60
DB 13 AVGPDVFAHQEDTERVLTNLNIGAEELRDPGLGAQFRVHLVKMVLTEPEGAPNITAN 72
OY 61 LTSSLLSVCGWSQTNIPEDDTPGHADLVLYITRFDLELDPGNRQVRGVTLGGACSPW 120
DB 73 LTSSLLSVCGWSQTNIPEDDTPGHADLVLYITRFDLELDPGNRQVRGVTLGGACSPW 132
OY 121 SCLITEDTGFDLGVTI 136
DB 133 SCLITEDTGFDLGVTI 148
RESULT 5
ABE04153
ID ABB04153 standard; protein; 1416 AA.
XX
XX ABB04153;
XX
XX 26-MAR-2002 (first entry)
XX
XX Human ADAMTS-M polypeptide.
XX
XX Osteoarthritis; rheumatoid arthritis; inflammatory bowel disease;
XX Crohn's disease; asthma; Alzheimer's disease; organ transplant rejection;
XX cachexia; allergy; cancer; leukaemia; lymphoma; osteoporosis;
XX atherosclerosis; congestive heart failure; myocardial infarction; stroke;
XX neurodegenerative disease; autoimmune disorder; Huntington's;

```

KW	Parkinson's; migraine; pain; depression; multiple sclerosis; burn;	
KW	infertility; diabetic shock; gene therapy; ADAMTS-M;	
KW	A Disintegrin And Metalloprotease; thrombospondin domain.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..97
FT		/label= Prodomain
FT		/note= "The mature form of the ADAMTS-M protein is
FT		processed by furin cleavage of the prodomain"
FT	Cleavage-site	94..97
FT		/label= Furin_cleavage_site
FT	Protein	98..1416
FT		/label= Mature ADAMTS-M protein
FT		/note= "The mature form of the ADAMTS-M protein is
FT		processed by furin cleavage of the prodomain"
FT	Domain	98..311
FT		/label= Metalloprotease_domain
FT	Domain	247..272
FT		/label= Zinc-binding_motif
FT	Domain	324..394
FT		/label= Disintegrin_domain
FT	Domain	410..473
FT		/label= Thrombospondin_submotif
FT	Domain	419..424
FT		/label= Heparin-binding_domain
FT	Domain	1099..1156
FT		/label= Thrombospondin_submotif
XX		
PN	EP1152055-A1.	
XX		
PD	07-NOV-2001.	
XX		
PF	24-APR-2001; 2001EP-0303706.	
XX		
PR	27-APR-2000; 2000US-200040P.	
XX		
PA	(PF12) PFIZER PROD INC.	
XX		
PI	Buckbinder L, Mitchell PG, Wachtmann TS, Walsh RT;	
XX		
DR	WPI: 2002-084275/12.	
DR	N-PSDB; ABA02549.	
XX		
PT	New polynucleotide, useful in gene therapy, particularly for treating	
PT	or preventing e.g. arthritis, Crohn's disease, Alzheimer's disease and	
PT	organ transplant toxicity and rejection, comprises ADAMTS	
PT	polynucleotide and encoded polypeptide -	
XX		
PS	Claim 4; Fig 2; 3lpp; English.	
XX		
CC	The present sequence represents a ADAMTS protein, designated ADAMTS-M,	
CC	that exhibits the characteristics of the ADAM (A Disintegrin And	
CC	Metalloprotease) family of metalloproteases, and contains a	
CC	thrombospondin domain (TS). The protein is encoded by the cDNA given in	
CC	ABA02549. The specification describes a newly isolated polynucleotide,	
CC	comprising a nucleotide sequence encoding an ADAMTS-M polypeptide as	
CC	given in the specification, or a metalloproteinase, disintegrin domain,	
CC	prodomain or its thrombospondin submotif. The polynucleotide, polypeptide	
CC	and agent are useful for manufacturing a medicament for treating a	
CC	subject in need of altering activity or expression of ADAMTS-M. The	
CC	polynucleotide, ADAMTS-M polypeptide and agent are useful for	
CC	manufacturing a medicament for treating arthritis (osteoarthritis and	
CC	rheumatoid arthritis), inflammatory bowel disease, Crohn's disease,	
CC	scleroderma, Alzheimer's disease, organ transplant toxicity and rejection,	
CC	cachexia, allergy, cancer (e.g. solid tumour cancer including colon,	
CC	breast, lung, prostate, brain or haematopoietic malignancies including	
CC	leukaemia and lymphoma), osteoporosis, atherosclerosis, aortic aneurysm,	
CC	congestive heart failure, myocardial infarction, stroke, head trauma,	
CC	spinal cord injury, neurodegenerative disease, autoimmune disorders,	
CC	Huntington's disease, Parkinson's disease, migraine, pain, depression,	
CC	multiple sclerosis, abnormal wound healing, burns, infertility or	

CC	diabetic shock. The polynucleotide and polypeptide are also useful for	
CC	diagnosing the diseases above. The polynucleotide is particularly useful	
CC	in gene therapy for treating the diseases cited above.	
XX		
SQ	Sequence	1416 AA;
	Query Match	100.0%; Score 716; DB 23; Length 1416;
	Best Local Similarity	100.0%; Pred. No. 8.1e-77;
	Matches 136; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1	AVGPDVFOAHOEDTERVLTNLNIGALLRDPSPSLGAQFRVHLVKMVLTPGEGAPNITAN 60
DB	110	AVGPDVFOAHOEDTERVLTNLNIGALLRDPSPSLGAQFRVHLVKMVLTPGEGAPNITAN 169
OY	61	LTSSLLSVCCWSQINPEDDTPGHADLVLYITRFDELDPGNRQVRGVLTOLGCACSPTW 120
DB	170	LTSSLLSVCCWSQINPEDDTPGHADLVLYITRFDELDPGNRQVRGVLTOLGCACSPTW 229
OY	121	SCLITEDTGFDLGVTI 136
DB	230	SCLITEDTGFDLGVTI 245
	RESULT 6	
	AAE24450	
ID	AAE24450 standard; Protein; 1427 AA.	
XX		
AC	AAE24450;	
XX		
DT	04-OCT-2002 (first entry)	
XX		
DE	Human Von Willebrand factor-cleaving protease (vWF-cp).	
XX		
KW	Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;	
KW	transgenic animal; immunisation; thromboembolic disease; pre-eclampsia;	
KW	thrombotic thrombocytopenic purpura; TTP; Henoch-Schonlein purpura;	
KW	thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;	
KW	transgenic; anticoagulant; chromosome 9.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..51
FT		/label= Signal_peptide
FT	Protein	52..1427
FT		/note= "Mature human vWF-cp protein"
FT	Cleavage-site	69..75
FT		/note= "Furin cleavage site"
FT	Region	224..228
FT		/note= "Catalytical side"
FT	Region	249
FT		/note= "Met turn"
FT	Domain	301..377
FT		/note= "Disintegrin like motif"
FT	Domain	387..439
FT		/note= "Thrombospondin type I motif"
FT	Region	441..553
FT		/note= "Cysteine rich region"
FT	Region	554..687
FT		/note= "Spacer"
FT	Domain	688..743
FT		/note= "Thrombospondin type I motif"
FT	Domain	744..805
FT		/note= "Thrombospondin type I motif"
FT	Domain	897..952
FT		/note= "Thrombospondin type I motif"
FT	Domain	953..1013
FT		/note= "Thrombospondin type I motif"
FT	Domain	1016..1073
FT		/note= "Thrombospondin type I motif"
FT	Domain	1075..1131
FT		/note= "Thrombospondin type I motif"
XX		

PN WO200242441-A2.
XX 30-MAY-2002.
XX 20-NOV-2001; 2001WO-EP13391.
XX 22-NOV-2000; 2000US-0721254.
XX 12-APR-2001; 2001US-0833328.
XX (BAXT) BAXTER AG.
XX Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
PI Scheiflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
PI Zimmermann K, Voelkel D;
XX WPI: 2002-479950/51.
XX N-PSDB; RAD39332.
XX Novel isolated or substantially purified Von Willebrand factor-cleaving
PT protease, useful for producing preparation for therapy of thrombosis
PT and thromboembolic disease such as thrombotic thrombocytic purpura
XX
XX Claim 1; Fig 5; 93pp; English.
XX The invention relates to an isolated or substantially pure Von Willebrand
CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
CC solution comprising vWF with the polypeptide ligand under conditions
CC where vWF is bound to the ligand and recovering from the ligand purified
CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
CC which involves immunising an animal with vWF-cp and isolating the anti-
CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
CC producing a preparation of prophylaxis and therapy of thrombosis and
CC thromboembolic disease such as thrombotic thrombocytic purpura (TTP).
CC Hemocho-Schonlein purpura, preeclampsia, neonatal thrombocytopenia or
CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
CC plasmatic or recombinantly produced vWF. The invention is useful for
CC construction expression systems and generating transgenic animals which
CC express the polypeptide in vivo. The present sequence is human vWF-cp
CC protein. vWF-cp gene is located on chromosome 9.
XX
SQ Sequence 1427 AA;
Query Match 100.0%; Score 716; DB 23; Length 1427;
Best Local Similarity 100.0%; Pred. No. 8.2e-77;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AVGPDVFOAHQEDTERVYLTNLNIGALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 60
Db 87 AVGPDVFOAHQEDTERVYLTNLNIGALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 146
Qy 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQRQVRGVTLGGACSPW 120
Db 147 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQRQVRGVTLGGACSPW 206
Qy 121 SCLITEDTGFDLGVTI 136
Db 207 SCLITEDTGFDLGVTI 222
RESULT 7
AAG63826
ID AAG63826 standard; Protein; 203 AA.
XX AAG63826;
XX 29-OCT-2001 (first entry)
DT
XX Amino acid sequence of a human zdtnt5 polypeptide.
XX Human; zdtnt5; anti-angiogenic; intestinal polypeptide; wound healing;
KW extracellular matrix interaction; tumour suppression; Gamete maturation;
KW immunologic recognition; gastrointestinal irradiation; chemotherapy;

KW proteolysis; apoptosis; angiogenesis; infection; cell adhesion;
KW cell fusion; cell signalling; tumour; Crohn's disease; melanoma; trauma;
KW inflammatory bowel disease; food poisoning; degenerative disease;
KW inflammation; fertility; gamete maturation; epithelial disorder.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Modified-site 69
FT FT /note= "potential N-linked glycosylation site"
FT Modified-site 73
FT FT /note= "potential N-linked glycosylation site"
FT Region 151..161
FT FT /note= "zinc-binding motif"
XX
XX WO200159112-A1.
XX 16-AUG-2001.
XX 09-FEB-2001; 2001WO-US04198.
XX 10-FEB-2000; 2000US-0501806.
XX (ZYMO) ZYMOGENETICS INC.
XX Holloway JL, Sheppard PO;
XX WPI: 2001-522477/57.
XX N-PSDB; AAH74759.
XX New anti-angiogenic intestinal polypeptides, zdtnt5 polypeptides, which
PT are members of disintegrin proteases, for modulating extracellular
PT matrix interaction, tumour suppression and wound healing
XX
XX Claim 1; Page 73-74; 92pp; English.
XX The present sequence represents a human zdtnt5 polypeptide. The zdtnt5
CC polypeptide is an anti-angiogenic intestinal polypeptide. zdtnt5 is
CC used for modulating extracellular matrix interactions. zdtnt5
CC polypeptide is useful as a tool for identifying new family members
CC of polypeptides. zdtnt5 polynucleotides are useful as probes or primers
CC to clone 5' non-coding regions of zdtnt5 gene. zdtnt5 polypeptides are
CC used for tumour suppression, gamete maturation, immunologic recognition,
CC and growth and differentiation either working in isolation or in
CC conjunction with other molecules in colon, small intestine, fetal lung,
CC testis and B-cells. zdtnt5 polypeptides are also useful for promoting
CC wound healing, in the treatment of disorders associated with recovery
CC after gastrointestinal irradiation, chemotherapy or antibody use
CC as anti-infectives, and extracellular matrix repair and remodeling. The
CC polypeptides are also useful for modulating proteolysis-apoptosis,
CC angiogenesis, infection, cell adhesion, cell fusion and signalling.
CC The polypeptides are also useful for treating tumour formation, Crohn's
CC disease, inflammatory bowel disease, food poisoning, melanoma,
CC degenerative diseases, disorders related to immunity, inflammation,
CC fertility, gamete maturation, immunology, trauma and epithelial
CC disorders.
XX
SQ Sequence 203 AA;
Query Match 99.6%; Score 713; DB 22; Length 203;
Best Local Similarity 99.3%; Pred. No. 1.1e-77;
Matches 135; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AVGPDVFOAHQEDTERVYLTNLNIGALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 60
Db 14 AVGPDVFOAHQEDTERVYLTNLNIGALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 73
Qy 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQRQVRGVTLGGACSPW 120
Db 74 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQRQVRGVTLGGACSPW 133
Qy 121 SCLITEDTGFDLGVTI 136
|||||

Dd 134 SCLITDGTGFDLGVTI 149

RESULT 8

AA663829

ID AA663829 standard; Protein; 1120 AA.

XX AC

XX AA663829;

XX 29-OCT-2001 (first entry)

XX Amino acid sequence of a human zdint5 polypeptide.

XX Human; zdint5; anti-angiogenic; intestinal polypeptide; wound healing;

KW extracellular matrix interaction; tumour suppression; gamete maturation;

KW immunologic recognition; gastrointestinal irradiation; chemotherapy;

KW proteolysis; apoptosis; angiogenesis; infection; cell adhesion;

KW cell fusion; cell signalling; tumour; Crohn's disease; melanoma; trauma;

KW inflammatory bowel disease; food poisoning; degenerative disease;

KW inflammation; fertility; gamete maturation; epithelial disorder.

XX OS

XX Homo sapiens.

XX Location/Qualifiers

FT Modified-site 172 /note= "potential N-linked glycosylation site"

FT Modified-site 176 /note= "potential N-linked glycosylation site"

FT Modified-site 485 /note= "potential N-linked glycosylation site"

FT Modified-site 533 /note= "potential N-linked glycosylation site"

FT Modified-site 560 /note= "potential N-linked glycosylation site"

FT Modified-site 595 /note= "potential N-linked glycosylation site"

FT Modified-site 635 /note= "potential N-linked glycosylation site"

FT Misc-difference 474 /note= "unspecified residue encoded by TTN"

XX WO200159112-A1.

XX 16-AUG-2001.

XX 09-FEB-2001; 2001WO-US04198.

XX 10-FEB-2000; 2000US-0501806.

XX (ZYMO) ZYMOGENETICS INC.

XX Holloway JL, Sheppard PO;

XX WPI; 2001-522477/57.

XX N-PSDB; AAH74765.

XX New anti-angiogenic intestinal polypeptides, zdint5 polypeptides, which

PT are members of disintegrin proteases, for modulating extracellular

PT matrix interaction, tumour suppression and wound healing

XX

PS Claim 2; Page 84-88; 92pp; English.

XX The present sequence represents a human zdint5 polypeptide. The zdint5

CC polypeptide is an anti-angiogenic intestinal polypeptide. Zdint5 is

CC used for modulating extracellular matrix interactions. Zdint5

CC polypeptide is useful as a tool for identifying new family members

CC of polypeptides. Zdint5 polynucleotides are useful as probes or primers

CC to clone 5' non-coding regions of zdint5 gene. Zdint5 polypeptides are

CC used for tumour suppression, gamete maturation, immunologic recognition,

CC and growth and differentiation either working in isolation or in

CC conjunction with other molecules in colon, small intestine, fetal lung,

CC testis and B-cells. zdint5 polypeptides are also useful for promoting

CC wound healing, in the treatment of disorders associated with recovery

CC after gastrointestinal irradiation, chemotherapy or antibody use,

CC as anti-infectives, and extracellular matrix repair and remodeling. The

CC polypeptides are also useful for modulating proteolysis, apoptosis,

CC angiogenesis, infection, cell adhesion, cell fusion and signalling.

CC The polypeptides are also useful for treating tumour formation, Crohn's

CC disease, inflammatory bowel disease, food poisoning, melanoma,

CC degenerative diseases, disorders related to immunity, inflammation,

CC fertility, gamete maturation, immunology, trauma and epithelial

CC disorders.

XX SQ

XX Sequence 1120 AA;

Query Match 99.6%; Score 713; DB 22; Length 1120;

Best Local Similarity 99.3%; Pred. No. 1.3e-76;

Matches 135; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AVGPDPVQAOHEDTERVLTNLNIGALLRDPSPGAFRHLVVKWVLTPEGAPNITAN 60

Dd 117 AVGPDPVQAOHEDTERVLTNLNIGALLRDPSPGAFRHLVVKWVLTPEGAPNITAN 176

Oy 61 LTSLLSVCGWSQINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTLGGACSPW 120

Dd 177 LTSLLSVCGWSQINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTLGGACSPW 236

Oy 121 SCLITDGTGFDLGVTI 136

Dd 237 SCLITDGTGFDLGVTI 252

RESULT 9

AA663829

ID AA663829 standard; Protein; 118 AA.

XX AC

XX AA663829;

XX 08-FEB-2001 (first entry)

XX Human ORFX ORF2432 polypeptide sequence SEQ ID NO:4864.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;

KW vulnery; antipsoriatic; antiparkinsonian; nontropic; neuroprotective;

KW immunostimulant; osteopathic; antichratic; immunosuppressant; cardiac;

KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

KW hypotensive; dermatological; immunosuppressive; antiinflammatory;

KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;

KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

KW cholesterol ester storage; systemic lupus erythematosus; infection;

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antinflammatory disease; coagulation;

KW thrombosis; contraceptive.

XX OS

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shinkets RA, Leach M;

XX WPI; 2000-602362/57.

DR N-PSDB; AAC76877.

```

XX  Novel nucleic acids and peptides derived from open reading frame X.
PT  useful for treating e.g. cancers, proliferative disorders,
PT  neurodegenerative disorders and cardiovascular disease -
XX
XX  Claim 11; Page 4050; 5507pp; English.
XX
XX  AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC  which represent the human ORFX open reading frames 1 to 161. The ORFX
CC  sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC  antiproliferative; antiparkinsonian; neurotropic; immunosuppressive;
CC  osteoparalytic; anticonvulsant; antiarthritic; coagulant; vasotropic;
CC  immunostimulant; cardiatic; dermatolytic; coagulant; vasotropic;
CC  antidiabetic; hypotensive; dermatological; immunosuppressive;
CC  antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC  antithyroid; and antianemic. The sequences can be used for determining
CC  the presence of or predisposition to, or preventing or treating
CC  pathological conditions associated with an ORFX protein in gene therapy.
CC  Nucleic acids can be used to express ORFX proteins in gene therapy
CC  vectors. The proteins and nucleic acids may be used to treat cancers,
CC  proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC  graft vs host disease, cardiovascular disease, diabetes mellitus,
CC  hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC  erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC  bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC  allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC  nocturnal haemoglobinuria, antiinflammatory disease, to enhance
CC  coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX  Sequence 118 AA;
SQ
    Query Match      73.9%; Score 529; DB 21; Length 118;
    Best Local Similarity 97.1%; Pred. No. 8.8e-56;
    Matches 100; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
    QY 34 LGAGFRVHLVKMVLTEPEGAPNITANLTSSLLSVCGWSOTINPEDDTPGHADLVLYIT 93
    DB 1 LGAGFRVHLVKMVLTEPEGAPNITANLTSSLLSVCGWSOTINPEDDTPGHADLVLYIT 60
    QY 94 RFDELDPGNGRQVGVTLGGACSPWCLITEDTGFGLGVTI 136
    DB 61 RFDELDPGNGXAVRGVTLGGACSPWCLITEDTGFGLGVTI 103
    RESULT 10
    AAB72283
    ID AAB72283 standard; Protein; 997 AA.
    AC AAB72283;
    DT 14-MAY-2001 (first entry)
    DE Human ADAMTS-7 amino acid sequence.
    KW ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;
    KW tumour cachexia; inflammation; dermatoparaxis; EDS-VIIC; angiogenesis;
    KW Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;
    KW metastasis; embryogenesis; egg implantation; chromosome 15; ADAMTS-7.
    XX Homo sapiens.
    OS
    XX WO200111074-A2.
    PN
    XX 15-FEB-2001.
    PD
    XX 03-AUG-2000; 2000WO-US21223.
    PF
    XX 06-AUG-1999; 99US-0369364.
    PR
    XX (CLEV-) CLEVELAND CLINIC FOUND.
    PA (APTE/) APTE S. S.
    PA (HURS/) HURSKAINEN T. L.
    PA (HIRO/) HIROHATA S.

```

```

XX  Apte SS, Hurskainen TL, Hirohata S;
XX
XX  WPI; 2001-159978/16.
DR  N-PSDB; AAF63440.
XX
XX  Murine and human 'A Disintegrin-like And Metalloprotease domain with
PT  Thrombospondin type I motifs' proteins and the nucleic acids encoding
PT  them, useful for treating e.g. tumours, inflammation and arthritis -
XX
XX  Claim 15; Fig 4; 181pp; English.
XX
XX  This invention relates to murine and human ADAMTS-N (A disintegrin-like
CC  and metalloprotease domain with thrombospondin type I motifs) proteins,
CC  designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the
CC  invention are cDNA sequences encoding the proteins, and antibodies
CC  specific for the proteins. The nucleic acid sequences and proteins may be
CC  used in the prevention, diagnosis and treatment of diseases associated
CC  with inappropriate ADAMTS-N expression. Disorders that may be treated
CC  using the nucleic acids, proteins and antibodies include, for example
CC  tumour cachexia, inflammation, dermatoparaxis in cattle or Ehlers-Danlos
CC  syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage
CC  in arthritic (both inflammatory and non-inflammatory) disease,
CC  angiogenesis, tumour growth and metastases, and they may also be used for
CC  controlling embryogenesis and implantation of fertilised eggs. The
CC  present sequence represents human ADAMTS-7. The human ADAMTS-7 gene is
XX  located on chromosome 15.
XX  Sequence 997 AA;
SQ
    Query Match      29.7%; Score 212.5; DB 22; Length 997;
    Best Local Similarity 37.5%; Pred. No. 3.7e-16;
    Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;
    QY 11 QEDTRYVLTNLNIGAEILLRDLGAGQFRVHLVKMVLTEPEGAPNITANLTSSLLSVCG 70
    DB 260 QQVESYVLTINMVGAGLPHDPSIGNPHITIVRLVLEDEEDLKITHADNTLKSFC 319
    QY 71 WSGTINPEDDTPGHADLVLYITRFDELPLDGNR--QVRGVTLGGACSPWCLITEDT 128
    DB 320 WQKSINMKGDAHPLHDDTAILLTRKDL-CAAMNRPCTLGLSHVAGMCPHRSCTINEDT 378
    QY 129 GFDLGVTI 136
    DB 379 GLPLAFTV 386
    RESULT 11
    AAB74944
    ID AAB74944 standard; Protein; 1686 AA.
    AC AAB74944;
    DT 02-JUL-2001 (first entry)
    DE Human ADAM type metal protease MDTs1 protein SEQ ID NO:1.
    XX Human; a disintegrin and metalloprotease type metal protease; MDTs1;
    KW MDTs2; MDTs3; ADAM type metal protease; cytostatic; antiarthritic;
    KW cancer; arthritis; arthrosis deformans.
    XX Homo sapiens.
    OS
    XX JP2001008687-A.
    PN
    XX 16-JAN-2001.
    PD
    XX 25-JUN-1999; 99JP-0180973.
    PF
    XX 25-JUN-1999; 99JP-0180973.
    PR
    XX (YAMA ) YAMANOUCHI PHARM CO LTD.
    PA

```

DR WPI; 2001-285362/30.
XX N-PSDB; AAF82149.
XX New metal protease and metal protease gene, for use as a drug for
PT cancers, arthritis and arthrosis deformans
XX Claim 1; Page 12-17; 31pp; Japanese.
XX The present sequence represents a disintegrin and metalloprotease (ADAM)
CC type metal protease designated MDT51, isolated from human. MDT5 proteins
CC have cytostatic and antiarthritic activities. They can be used as a drug
CC for cancers, arthritis and arthrosis deformans.
XX
SQ Sequence 1686 AA;
Query Match 29.7%; Score 212.5; DB 22; Length 1686;
Best Local Similarity 37.5%; Pred. No. 7.9e-16;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;
Qy 11 QEDTERYVLTNIGAEILLRPSLGAQFRVHLVQVILTEGAPNITANLTSSILSYCG 70
Db 260 QPQVESYVLTNIMVAGLFHDSIGNPIHITIVRLVLEDEEDLKITHADNTLKSPCK 319
Qy 71 WSQTINPEDDTPGHADLVLYITRDLPLDGNR--QVRGVTLGGACSPWMSCLITBDT 128
Db 320 WOKSINMGDAHLPHLDHDTAILTRKDL-CAANNRPCEITGLSHVAGMCPHRSINSNEDT 378
Qy 129 GPDGLGVTI 136
Db 379 GLPLAFTV 386
RESULT 12
AAE00913
ID AAE00913 standard; Protein; 1686 AA.
XX
AC AAE00913;
XX
DT 04-JUL-2001 (first entry)
XX
DE Human 27875 ADAM-TS protein, alternative version.
XX
KW Human; ADAM-TS; A Disintegrin And Metalloproteinase; antiinflammatory;
KW anglogenesis; bronchial asthma; Goodpasture's syndrome; metastasis;
KW heart failure; cardiac hypertrophy; chronic ischaemic heart disease;
KW sickle cell disease nephropathy; urinary tract obstruction; haenostatic;
KW skeletal muscle disorder; myocardial infarction; blood vessel disorder;
KW hypertension; atherosclerosis; vasculitis; renal artery stenosis; mumps;
KW gonorrhea; tuberculosis; syphilis; spermatocytic seminoma; osteoporosis;
KW rhabdomyosarcoma; glomerulonephritis; bone disorder; Paget's disease;
KW rickets; osteomalacia; Hodgkin's disease; gene therapy; antibacterial;
KW candidant; tumour; thymoma; vasotropic; cytostatic; virucide.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= Signal_peptide
FT Modified-site 6..8
FT /note= "Protein kinase C phosphorylation site"
FT Protein 31..1686
FT /label= Human_mature_27875_ADAM-TS_protein
FT Modified-site 55..60
FT /note= "N-myristoylation site"
FT Modified-site 73..75
FT /note= "Protein kinase C phosphorylation site"
FT Domain 78..93
FT /note= "Crystallins beta and gamma Greek key motif"
FT Modified-site 94..97
FT /note= "N-glycosylation site"
FT Modified-site 110..112
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 115..120

FT /note= "N-myristoylation site"
FT 141..146
FT /note= "N-myristoylation site"
FT 147..150
FT /note= "Casein kinase II phosphorylation site"
FT 159..162
FT /note= "Casein kinase II phosphorylation site"
FT 195..197
FT /note= "Cell attachment sequence"
FT 214..217
FT /note= "Casein kinase II phosphorylation site"
FT 214..216
FT /note= "Protein kinase C phosphorylation site"
FT 244..259
FT /label= Metalloproteinase_domain
FT 313..315
FT /note= "Protein kinase C phosphorylation site"
FT 342..345
FT /note= "Casein kinase II phosphorylation site"
FT 342..344
FT /note= "Protein kinase C phosphorylation site"
FT 373..376
FT /note= "Casein kinase II phosphorylation site"
FT 379..384
FT /note= "N-myristoylation site"
FT 385..394
FT /label= Zinc_binding_domain
FT 401..404
FT /note= "Casein kinase II phosphorylation site"
FT 408..411
FT /note= "Amidation site"
FT 479..484
FT /note= "N-myristoylation site"
FT 488..567
FT /label= Thrombospondin_domain
FT 505..508
FT /note= "Casein kinase II phosphorylation site"
FT 513..518
FT /note= "N-myristoylation site"
FT 539..545
FT /note= "Growth factor and cytokine metalloproteinase family signature 2 domain"
FT 539..544
FT /note= "N-myristoylation site"
FT 541..592
FT /label= Disintegrin_domain
FT 542..592
FT /label= Thrombospondin_domain
FT 557..562
FT /note= "N-myristoylation site"
FT 569..571
FT /note= "Protein kinase C phosphorylation site"
FT 596
FT /label= Unknown
FT /note= "Encoded by CRC"
FT 598..600
FT /note= "Protein kinase C phosphorylation site"
FT 605..608
FT /note= "Casein kinase II phosphorylation site"
FT 614..619
FT /note= "N-myristoylation site"
FT 667..672
FT /note= "N-myristoylation site"
FT 687..692
FT /note= "Cytochrome C family heme-binding site"
FT 688..693
FT /note= "N-myristoylation site"
FT 693..696
FT /note= "N-glycosylation site"
FT 703..706
FT /note= "Casein kinase II phosphorylation site"
FT 716..721
FT /note= "N-myristoylation site"

```

FT Modified-site 740..747 /note= "Tyrosine kinase phosphorylation site"
FT Modified-site 765..770 /note= "N-myristoylation site"
FT Modified-site 774..779 /note= "N-myristoylation site"
FT Modified-site 778..781 /note= "N-glycosylation site"
FT Domain 825..868 /label= Thrombospondin_domain
FT Modified-site 872..875 /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 901..903 /note= "Protein kinase C phosphorylation site"
FT Modified-site 917..920 /note= "Casein kinase II phosphorylation site"
FT Domain 949..994 /label= Thrombospondin_domain
FT Domain 949..988 /label= Thrombospondin_domain
FT Modified-site 950..953 /note= "N-glycosylation site"
FT Modified-site 957..960 /note= "Casein kinase II phosphorylation site"
FT Modified-site 962..964 /note= "Protein kinase C phosphorylation site"
FT Modified-site 971..974 /note= "N-glycosylation site"
FT Modified-site 1005..1010 /note= "N-myristoylation site"
FT Binding-site 1006..1009 /note= "Glycosaminoglycan attachment site"
FT Modified-site 1011..1014 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1035..1037 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1039..1044 /note= "N-myristoylation site"
FT Modified-site 1192..1195 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1263..1252 /note= "N-myristoylation site"
FT Modified-site 1263..1268 /note= "N-myristoylation site"
FT Modified-site 1308..1311 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1358..1363 /note= "N-myristoylation site"
FT Modified-site 1370..1372 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1385..1387 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1397..1400 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1412..1415 /note= "N-glycosylation site"
FT Domain 1415..1463 /label= Thrombospondin_domain
FT Modified-site 1419..1422 /note= "N-glycosylation site"
FT Modified-site 1440..1443 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1440..1442 /note= "Protein kinase C phosphorylation site"
FT Domain 1466..1521 /label= Thrombospondin_domain
FT Modified-site 1470..1473 /note= "N-glycosylation site"
FT Modified-site 1483..1486 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1483..1485 /note= "Protein kinase C phosphorylation site"

```

```

FT Modified-site 1517..1522 /note= "N-myristoylation site"
FT Modified-site 1528..1531 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1528..1530 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1546..1549 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1592..1597 /note= "N-myristoylation site"
FT Modified-site 1599..1601 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1606..1609 /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"

Query Match 29.7%; Score 212.5; DB 22; Length 1686;
Best Local Similarity 37.5%; Pred. No 7.9e-16;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;

QY 11 QEDTERVLTNLTNIGAEILLRPSLQAFRVHLYKVMVLTEPEGAPNITANLTSSLVSCG 70
Db 260 QPQVESYVLTIMNVAGLFHDPISIGNPIHITIVRLVLEDEEDLKHTHADNTLSKFC 319
QY 71 WSGTIPEDDTPGHADLVLYTRFLELPDGNR--QVRGVTLGGACSTWKLITEDT 128
Db 320 WQKSINMKGAHPLHHDHTAILLTRKDL-CAAMNRPCTGLSLHVGKVCOPHRSCSINEDT 378
QY 129 GFDLGVTI 136
Db 379 GLPLAFTV 386

RESULT 13
AAE00934
ID AAE00934 standard; Protein; 1686 AA.
XX
AC AAE00934;
XX
DT 04-JUL-2001 (first entry)
XX
DE Human 27875 ADAM-TS (a disintegrin and metalloproteinase).
XX
KW Human; ADAM-TS; A Disintegrin And Metalloproteinase; antiinflammatory;
KW angio genesis; bronchial asthma; Goodpasture's syndrome; metastasis;
KW heart failure; cardiac hypertrophy; chronic ischaemic heart disease;
KW sickle cell disease nephropathy; urinary tract obstruction; haemostatic;
KW skeletal muscle disorder; myocardial infarction; blood vessel disorder;
KW hypertension; atherosclerosis; vasculitis; renal artery stenosis; mumps;
KW gonorrhea; tuberculosis; syphilis; spermatocytic seminoma; osteoporosis;
KW rhabdomyosarcoma; glomerulonephritis; bone disorder; Paget's disease;
KW rickets; osteomalacia; Hodgkin's disease; gene therapy; antibacterial;
KW cardiac; tumour; thymoma; vasotropic; cytostatic; virucide.
XX
OS Homo sapiens.
XX
FH Key
FH Peptide
FT /label= Signal_peptide
FT Modified-site 6..8
FT Protein 31..1686
FT Modified-site 55..60 /label= Human_mature_27875_ADAM-TS_protein
FT Modified-site 73..75 /note= "N-myristoylation site"
FT Domain 78..93 /note= "Protein kinase C phosphorylation site"
FT Modified-site 94..97 /note= "Crystallins beta and gamma Greek key motif"
FT Modified-site 110..112 /note= "N-glycosylation site"
FT /note= "Protein kinase C phosphorylation site"

```

FT Modified-site 115..120 /note= "N-myristoylation site"
FT Modified-site 141..146 /note= "N-myristoylation site"
FT Modified-site 147..150 /note= "Casein kinase II phosphorylation site"
FT Modified-site 159..162 /note= "Casein kinase II phosphorylation site"
FT Binding-site 195..197 /note= "Cell attachment sequence"
FT Modified-site 214..217 /note= "Casein kinase II phosphorylation site"
FT Modified-site 214..216 /note= "Protein kinase C phosphorylation site"
FT Domain 244..259 /label= Metalloproteinase_domain
FT Modified-site 313..315 /note= "Protein kinase C phosphorylation site"
FT Modified-site 342..345 /note= "Casein kinase II phosphorylation site"
FT Modified-site 342..344 /note= "Protein kinase C phosphorylation site"
FT Modified-site 373..376 /note= "Casein kinase II phosphorylation site"
FT Modified-site 379..384 /note= "N-myristoylation site"
FT Domain 385..394 /label= Zinc_binding_domain
FT Modified-site 401..404 /note= "Casein kinase II phosphorylation site"
FT Modified-site 408..411 /note= "Amidation site"
FT Modified-site 479..484 /note= "N-myristoylation site"
FT Domain 488..567 /label= Thrombospondin_domain
FT Modified-site 505..508 /note= "Casein kinase II phosphorylation site"
FT Modified-site 513..518 /note= "N-myristoylation site"
FT Domain 539..545 /note= "Growth factor and cytokine metalloproteinase family signature 2 domain"
FT Modified-site 539..544 /note= "N-myristoylation site"
FT Domain 541..592 /label= Disintegrin_domain
FT Domain 542..592 /label= Thrombospondin_domain
FT Modified-site 557..562 /note= "N-myristoylation site"
FT Modified-site 569..571 /note= "Protein kinase C phosphorylation site"
FT Modified-site 598..600 /note= "Protein kinase C phosphorylation site"
FT Modified-site 605..608 /note= "Casein kinase II phosphorylation site"
FT Modified-site 614..619 /note= "N-myristoylation site"
FT Modified-site 667..672 /note= "N-myristoylation site"
FT Binding-site 687..692 /note= "Cytochrome C family heme-binding site"
FT Modified-site 688..693 /note= "N-myristoylation site"
FT Modified-site 693..696 /note= "N-glycosylation site"
FT Modified-site 703..706 /note= "Casein kinase II phosphorylation site"
FT Modified-site 716..721 /note= "N-myristoylation site"
FT Modified-site 740..747 /note= "Tyrosine kinase phosphorylation site"

FT Modified-site 765..770 /note= "N-myristoylation site"
FT Modified-site 774..779 /note= "N-myristoylation site"
FT Modified-site 778..781 /note= "N-glycosylation site"
FT Domain 825..868 /label= Thrombospondin_domain
FT Modified-site 872..875 /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 901..903 /note= "Protein kinase C phosphorylation site"
FT Modified-site 917..920 /note= "Casein kinase II phosphorylation site"
FT Domain 949..994 /label= Thrombospondin_domain
FT Domain 949..988 /label= Thrombospondin_domain
FT Modified-site 950..953 /note= "N-glycosylation site"
FT Modified-site 957..960 /note= "Casein kinase II phosphorylation site"
FT Modified-site 962..964 /note= "Protein kinase C phosphorylation site"
FT Modified-site 971..974 /note= "N-glycosylation site"
FT Modified-site 1005..1010 /note= "N-myristoylation site"
FT Binding-site 1006..1009 /note= "Glycosaminoglycan attachment site"
FT Modified-site 1011..1014 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1035..1037 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1039..1044 /note= "N-myristoylation site"
FT Modified-site 1192..1195 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1263..1252 /note= "N-myristoylation site"
FT Modified-site 1263..1268 /note= "N-myristoylation site"
FT Modified-site 1308..1311 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1358..1363 /note= "N-myristoylation site"
FT Modified-site 1370..1372 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1385..1387 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1397..1400 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1412..1415 /note= "N-glycosylation site"
FT Domain 1415..1463 /label= Thrombospondin_domain
FT Modified-site 1419..1422 /note= "N-glycosylation site"
FT Modified-site 1440..1443 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1440..1442 /note= "Protein kinase C phosphorylation site"
FT Domain 1466..1521 /label= Thrombospondin_domain
FT Modified-site 1470..1473 /note= "N-glycosylation site"
FT Modified-site 1483..1486 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1483..1485 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1517..1522 /note= "N-myristoylation site"

```
FT Modified-site 1528..1531
FT /note="Casein kinase II phosphorylation site"
FT Modified-site 1528..1530
FT /note="Protein kinase C phosphorylation site"
FT Modified-site 1546..1549
FT /note="Casein kinase II phosphorylation site"
FT Modified-site 1592..1597
FT /note="N-myristoylation site"
FT Modified-site 1599..1601
FT /note="Protein kinase C phosphorylation site"
FT Modified-site 1606..1609
FT /note="cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 1620..1622
FT /note="Protein kinase C phosphorylation site"
FT Modified-site 1625..1630

Query Match 29.7%; Score 212.5; DB 22; Length 1686;
Best Local Similarity 37.5%; Pred. NO. 7.9e-16;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;

QY 11 QEDTERVLTNLNIGALLRDPGLGQAFRVHLVKVILTEPEGAPNITANLTSSLSVCG 70
DB 260 QPQVESYVLTIMNMVAGLFPHDPSIGNPIHITIVRLVLEDEEDLKITHADNTLKSFC 319
QY 71 WSGTINPEDDTPGHADLVLYITRFDLELPDGNR--QVRGVLTOLGGACSPWTSCLITE 128
DB 320 WQKSINMKGDAPLHHDHTAILLTRKDL-CAAMNRPCTGLSHVAGMCQPHRSCSINED 378
QY 129 GFDLGVTI 136
DB 379 GLPLAFTV 386

RESULT 14
AAB41379
ID AAB41379 standard; Protein; 1784 AA.
XX
AC AAB41379;
XX
XX 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1143 polypeptide sequence SEQ ID NO:2286.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
PA
```

```
XX PI Shimkets RA, Leach M;
XX WPI; 2000-602362/57.
XX DR N-PSDB; AAC75588.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 1678-1682; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antithyroid; and antianaemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 1784 AA;
XX
Query Match 29.7%; Score 212.5; DB 21; Length 1784;
Best Local Similarity 37.5%; Pred. NO. 8.6e-16;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;

QY 11 QEDTERVLTNLNIGALLRDPGLGQAFRVHLVKVILTEPEGAPNITANLTSSLSVCG 70
DB 263 QPQVESYVLTIMNMVAGLFPHDPSIGNPIHITIVRLVLEDEEDLKITHADNTLKSFC 322
QY 71 WSGTINPEDDTPGHADLVLYITRFDLELPDGNR--QVRGVLTOLGGACSPWTSCLITE 128
DB 323 WQKSINMKGDAPLHHDHTAILLTRKDL-CAAMNRPCTGLSHVAGMCQPHRSCSINED 381
QY 129 GFDLGVTI 136
DB 382 GLPLAFTV 389

RESULT 15
AAU72897
ID AAU72897 standard; Protein; 1505 AA.
XX
XX AAU72897;
XX
XX 26-FEB-2002 (first entry)
XX
DE Human metalloprotease partial protein sequence #9.
XX
XX Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;
XX vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;
XX hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;
XX anorectic; antinflammatory; aspartyl protease; cysteine protease; colon;
XX metalloprotease; serine protease; cancer; haematopoietic; breast; lung;
XX prostate; cervical; brain; ovarian; bladder; kidney; pain;
XX immune-related disease; cardiovascular disease; neuronal disorder;
XX migraine; sexual dysfunction; mood disorder; attention disorder;
XX cognition disorder; hypotension; hypertension; psychotic disorder;
XX dyskinesia; metabolic disorder; inflammatory disorder.
KW
```

```
XX OS Homo sapiens.
XX PN WO200183782-A2.
XX PD 08-NOV-2001.
XX PF 04-MAY-2001; 2001WO-US14431.
XX PR 04-MAY-2000; 2000US-201879P.
XX PA (SUGEN-) SUGEN INC.
XX PI Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
XX PI Payne V;
XX DR WPI: 2002-041502/05.
XX DR N-PSDB; AAS97180.
XX PT Novel protease polypeptide useful for screening for substances that may
XX PT be used to treat, e.g., cancers, immune-related diseases,
XX PT cardiovascular disease, migraine, pain, psychotic and inflammatory
XX PT disorders -
XX PS Claim 28; Figure 2F; 232pp; English.
XX CC The invention relates to an isolated, enriched, or purified protease
XX CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
XX CC screen for substances (S) that may modulate its activity. Administering
XX CC S (which modulates protease activity in vitro) may be used to treat a
XX CC disease or disorder selected from cancers (e.g., of tissues, of blood or
XX CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,
XX CC brain, ovarian, bladder or kidney), immune-related diseases and
XX CC disorders, cardiovascular disease, brain or neuronal-associated diseases
XX CC (e.g., central or peripheral nervous system diseases, migraine, pain,
XX CC sexual dysfunction, mood disorders, attention disorders, cognition
XX CC disorders, hypotension, hypertension, psychotic disorders, neurological
XX CC disorders and dyskinesias), metabolic disorders and inflammatory
XX CC disorders. (I) may also be useful as a diagnostic tool for a disease or
XX CC disorder such as those above. AAU72876-AAU72910 represent human
XX CC protease amino acid sequences of the invention.
XX SQ Sequence 1505 AA;
Query Match 29.3%; Score 210; DB 23; Length 1505;
Best Local Similarity 34.4%; Pred. No. 1.3e-15;
Matches 45; Conservative 23; Mismatches 63; Indels 0; Gaps 0;
QY 6 VFQAHQEDTERVYLTNLNIGALLDDPSLGAQFRVHLVKNWVLTPEGAPNITANLTSSL 65
DB 271 VVSAHGSNLQNYLTILMSIVATIKDPSIGNLIHVVVKLVMIHREEGPVINFDGATIL 330
QY 66 LSVCGMSQITNPEDDTPGHADLVLYITRFDLELPGNRQVRGVTLQGCACSPWTWCSCLIT 125
DB 331 KNFCSWQQTQNDLDDVHPSSHHTAVLITREDICSSKEKCNMLGLSYLGTICDPLQSCFIN 390
QY 126 EDTGFDLGVTI 136
DB 391 EEKGLISAFIT 401
Search completed: March 20, 2003, 12:33:30
Job time : 75.7568 secs
```

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:31:32 ; Search time 4.37838 Seconds
(without alignments)
564.722 Million cell updates/sec

Title: SEQID_1

Perfect score: 57

Sequence: 1 AAGGILHLELLV 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	100.0	1427	Q96L37	Q96L37 homo sapien
2	42	73.7	355	Q9LWH0	Q9LWH0 oryza sativ
3	42	73.7	694	Q9Y4A8	Q9Y4A8 homo sapien
4	39	68.4	175	Q9BP09	Q9BP09 homo sapien
5	38	66.7	151	Q96A05	Q96A05 trypanosoma
6	38	66.7	316	Q9UC64	Q9UC64 homo sapien
7	38	66.7	359	Q96YK1	Q96YK1 sulfolobus
8	38	66.7	384	Q96CP5	Q96CP5 homo sapien
9	38	66.7	489	Q60416	Q60416 homo sapien
10	38	66.7	490	Q96FV4	Q96FV4 homo sapien
11	38	66.7	1081	Q68831	Q68831 bacteroides
12	37	64.9	116	Q9YAE8	Q9YAE8 aeropyrum p
13	37	64.9	322	Q93P08	Q93P08 prochloroco
14	37	64.9	322	Q93P05	Q93P05 prochloroco
15	37	64.9	322	Q93P04	Q93P04 prochloroco
16	37	64.9	352	Q07295	Q07295 prochloroco

17	37	64.9	356	17	Q97WN3	Q97WN3 sulfolobus
18	37	64.9	388	16	Q8XVW1	Q8XVW1 ralbstonia s
19	37	64.9	454	2	Q93P65	Q93P65 pseudomonas
20	37	64.9	477	13	Q8QH19	Q8QH19 oncorhynchus
21	37	64.9	555	10	Q9FI58	Q9FI58 arabidopsis
22	37	64.9	555	10	Q94BN5	Q94BN5 arabidopsis
23	37	64.9	576	2	Q8TR5	Q8TR5 uncultured
24	37	64.9	732	4	Q96T17	Q96T17 homo sapien
25	37	64.9	876	16	Q93IG6	Q93IG6 mycobacteri
26	37	64.9	1326	5	Q9W4N4	Q9W4N4 drosophila
27	37	64.9	1423	5	Q95YP5	Q95YP5 leishmania
28	36	63.2	233	16	Q86804	Q86804 streptomyce
29	36	63.2	335	12	Q8V6J1	Q8V6J1 trocara vir
30	36	63.2	335	12	Q8V6J0	Q8V6J0 trocara vir
31	36	63.2	451	5	Q95T70	Q95T70 drosophila
32	36	63.2	470	5	Q9VG87	Q9VG87 drosophila
33	36	63.2	480	4	Q96DD2	Q96DD2 homo sapien
34	36	63.2	505	16	Q8UAA5	Q8UAA5 agrobacteri
35	36	63.2	597	4	Q9NVR2	Q9NVR2 homo sapien
36	36	63.2	597	11	Q9DSU1	Q9DSU1 mus musculu
37	36	63.2	842	3	Q9P6K4	Q9P6K4 schizosacch
38	36	63.2	1142	10	Q40263	Q40263 mesotaenium
39	35	61.4	151	5	O18573	O18573 andrena eri
40	35	61.4	165	17	Q981B4	Q981B4 sulfolobus
41	35	61.4	170	10	Q94KG6	Q94KG6 phaseolus v
42	35	61.4	170	10	Q94KG3	Q94KG3 phaseolus v
43	35	61.4	200	2	Q8RPP4	Q8RPP4 legionella
44	35	61.4	218	16	Q9RDP9	Q9RDP9 streptomyce
45	35	61.4	231	10	Q94KP4	Q94KP4 phaseolus v

ALIGNMENTS

RESULT 1

Q96L37 PRELIMINARY; PRT; 1427 AA.
ID Q96L37
AC Q96L37; TREMBLrel. 19, Created
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Von Willebrand factor-cleaving protease precursor.
GN ADAMTS13.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

SEQUENCE FROM N.A.

RP TISSUE=LIVER;
RX PubMed=11557746;
RA Zheng X., Chung D., Takayama T.K., Majerus E.M., Sadler J.E.,
RA Fujikawa K.;
RT "Structure of von Willebrand Factor-cleaving Protease (ADAMTS13), a
RT Metalloprotease Involved in Thrombotic Thrombocytopenic Purpura.";
RL J. Biol. Chem. 276:41059-41063 (2001).
DR EMBL; AY055376; AAL17652.1; -

DR MEROPS; M12.241; -

DR InterPro; IPR001590; Reprolysin.

DR InterPro; IPR000884; TSF1.

DR InterPro; IPR000130; Zn.Mtpeptdse.

DR Pfam; PF01421; Reprolysin; 1.

DR Pfam; PF00090; tsf_1; 4.

DR PROSITE; PS50215; ADAM_MERO; 1.

DR PROSITE; PS50092; TSF1_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

Query Match 100.0%; Score 57; DB 4; Length 1427;
Best Local Similarity 100.0%; Pred. No. 0.04;


```

CC -1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-
CC THIOLESTER FORMATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
DR EMBL: AF284603; AAK82982.1; -.
DR InterPro: IPR000515; BPD.transp.
DR InterPro: IPR000608; UBQ.conjugat.
DR Pfam: PF001179; UQ.con; 1.
DR ProDom: PD000461; UBQ.conjugat; 1.
DR ProSITE: PS00402; BPD.TRANSF.INN.MEMBR.; UNKNOWN_1.
DR ProSITE: PS00183; UBIQUITIN.CONJUGAT_1; UNKNOWN_1.
DR ProSITE: PS0127; UBIQUITIN.CONJUGAT_2; 1.
DR KW Ligase; Ubiquitin conjugation.
SQ SEQUENCE 151 AA; 17473 MW; 11E72C9720E27748 CRC64;

Query Match 66.7%; Score 38; DB 5; Length 151;
Best Local Similarity 70.0%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGILHLELLV 12
DB 48 GGIFHLEFL 57

RESULT 6
Q9UG64 PRELIMINARY; PRT; 316 AA.
AC Q9UG64;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 35.2 kDa protein (fragment).
GN DKF258611223.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Anorg W., Winkner U., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL050216; CAB43319.1; -.
DR MEROPS: M16.975; -.
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 316 AA; 35188 MW; AD9FD73A65FA478B CRC64;

Query Match 66.7%; Score 38; DB 4; Length 316;
Best Local Similarity 72.7%; Pred. No. 39;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGILHLELL 11
DB 70 AAGGVSHDELL 80

RESULT 7
Q96YK1 PRELIMINARY; PRT; 359 AA.
AC Q96YK1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative aspartate-semialdehyde dehydrogenase.
GN ST2171.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Anai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kuenida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermosacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL: AP000989; BAB67276.1; -.
DR InterPro: IPR000534; Semialdh_dh.
DR Pfam: PF01118; Semialdehyde_dh; 1.
DR Pfam: PF02774; Semialdehyde_dhC; 1.
DR TIGRFAMs: TIGR00978; asd_EA; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 359 AA; 39559 MW; 426AAE51597DD0E8 CRC64;

Query Match 66.7%; Score 38; DB 17; Length 359;
Best Local Similarity 75.0%; Pred. No. 45;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGGILHLELLV 12
DB 341 AGGILAEELLV 352

RESULT 8
Q96CP5 PRELIMINARY; PRT; 384 AA.
AC Q96CP5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 43.0 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=B-CELL;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC014079; AAH14079.1; -.
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16; 1.
KW Hypothetical protein.
SQ SEQUENCE 384 AA; 43029 MW; 96FF3C1ABC7AB912 CRC64;

Query Match 66.7%; Score 38; DB 4; Length 384;
Best Local Similarity 72.7%; Pred. No. 48;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGILHLELL 11
DB 138 AAGGVSHDELL 148

RESULT 9
Q60416 PRELIMINARY; PRT; 489 AA.
AC Q60416;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE WUGSC:H.RG276003.2 protein.
GN WUGSC:H.RG276003.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

RN SEQUENCE FROM N.A.
 RA Rohlfing T., Wohldmann P., Antoniou B., Bauer C., O'Neal D.;
 RL "The sequence of Homo sapiens BAC clone CTA-27603."
 RT Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC004668; AAC15866.1; -.
 DR MEROPS; M16.003; -.
 DR InterPro; IPR001431; Peptidase M16.
 DR Pfam; PF00675; Peptidase_M16; 1.
 DR PROSITE; PS00143; INSULINASE; 1.
 SQ SEQUENCE 489 AA; 54366 MW; 79250D016E60CFEE CRC64;

Query Match 66.7%; Score 38; DB 4; Length 489;
 Best Local Similarity 72.7%; Pred. No. 62;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
 ||||:||||
 Db 243 AAGGVSHDELL 253

RESULT 10

Q96FV4 PRELIMINARY; PRT; 490 AA.
 AC Q96FV4;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Unknown (protein for MGC:13691).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN SEQUENCE FROM N.A.
 RA TISSUE=SKETAL MUSCLE;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC010398; AAH10398.1; -.
 DR MEROPS; M16.975; -.
 DR InterPro; IPR001431; Peptidase M16.
 DR Pfam; PF00675; Peptidase_M16; 1.
 DR PROSITE; PS00143; INSULINASE; UNKNOWN 1.
 SQ SEQUENCE 490 AA; 54191 MW; 8999F3F57A4D14E1 CRC64;

Query Match 66.7%; Score 38; DB 4; Length 490;
 Best Local Similarity 72.7%; Pred. No. 62;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
 ||||:||||
 Db 243 AAGGVSHDELL 253

RESULT 11

O68831 PRELIMINARY; PRT; 1081 AA.
 AC O68831;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Surface antigen BspA.
 GN BSPA.
 OS Bacteroides forsythus.
 OC Bacteria; CF8 group; Bacteroidetes; Bacteroidales; Bacteroidaceae;
 OC Bacteroides.
 OX NCBI_TaxID=28112;
 RN SEQUENCE FROM N.A.
 RP

RC STRAIN=ATCC43037;
 RX MEDLINE=99043895; PubMed=9826345;
 RA Sharma A., Sojar H.T., Glurich I., Honma K., Kuramitsu H.K.,
 RA Genco R.J.;
 RT "Cloning, expression, and sequencing of a cell surface antigen
 RT containing a leucine-rich repeat motif from Bacteroides forsythus ATCC
 RT 43037."
 RL Infect. Immun. 66:5703-5710(1998).
 DR EMBL; AF054892; AAC82625.1; -.
 DR InterPro; IPR003443; Big_2.
 DR Pfam; PF02368; Big_2; 1.
 DR PRINTS; PR01611; LIMPII.
 SQ SEQUENCE 1081 AA; 113908 MW; 6E5E36FA0ACDBDSC CRC64;

Query Match 66.7%; Score 38; DB 2; Length 1081;
 Best Local Similarity 80.0%; Pred. No. 1.4e+02;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLEL 10
 ||||:||||
 Db 1023 AAGGALHLL 1032

RESULT 12

Q9YAE8 PRELIMINARY; PRT; 116 AA.
 AC Q9YAE8;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein APE1991.
 GN APE1991.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
 OC Desulfurococaceae; Aeropyrum.
 OX NCBI_TaxID=56636;
 RN SEQUENCE FROM N.A.
 RP STRAIN=KJ;
 RC MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1."
 RL DNA Res. 6:83-101(1999).
 DR EMBL; AP000063; BAA81001.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 116 AA; 12506 MW; 2ADE1A5B2B82FA8A CRC64;

Query Match 64.9%; Score 37; DB 17; Length 116;
 Best Local Similarity 77.8%; Pred. No. 21;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGGILHLE 9
 ||||:||||
 Db 44 AAGGLQHLE 52

RESULT 13

Q93P06 PRELIMINARY; PRT; 322 AA.
 ID Q93P06;
 AC Q93P06;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Chlorophyll a/b-binding light-harvesting protein Pcb
 DE (Fragment).
 GN PCB.
 OS Prochlorococcus marinus.

OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
 OC Prochlorococcus.
 OX NCBI_TaxID=1219;
 RN [1]
 RC STRAIN=SB;
 RA Garczarek L., van der Staay G.W.M., Hess W.R., Le Gall F.,
 RA Partensky F.;
 RT "Expression and phylogeny of the multiple antenna genes of the low-
 RT light adapted strain Prochlorococcus marinus SS120
 RT (Oxyphotobacteria).";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF354645; AAK69280.1; -;
 DR InterPro; IPR000932; PSIIProt.
 DR InterPro; IPR000408; Reg_chr_condens.
 DR Pfam; PF00421; PSII; 1.
 DR PROSITE; PS00626; RCC1_2; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 322 322
 FT NON_TER 322 322
 SQ SEQUENCE 322 AA; 35495 MW; B0762A1DD6279032 CRC64;

Query Match 64.9%; Score 37; DB 2; Length 322;
 Best Local Similarity 72.7%; Pred. No. 62;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAGGILHLELL 11
 ||||:||||
 Db 98 AAGGLLHSLLL 108

RESULT 14

O93P05 PRELIMINARY; PRT; 322 AA.
 AC O93P05;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Chlorophyll a/b-binding light-harvesting protein Pcb
 DE (Fragment).
 GN PCB.
 OS Prochlorococcus marinus.
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
 OC Prochlorococcus.
 OX NCBI_TaxID=1219;
 RN [1]
 RC STRAIN=GP2;
 RA Garczarek L., van der Staay G.W.M., Hess W.R., Le Gall F.,
 RA Partensky F.;
 RT "Expression and phylogeny of the multiple antenna genes of the low-
 RT light adapted strain Prochlorococcus marinus SS120
 RT (Oxyphotobacteria).";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF354646; AAK69280.1; -;
 DR InterPro; IPR000932; PSIIProt.
 DR InterPro; IPR000408; Reg_chr_condens.
 DR Pfam; PF00421; PSII; 1.
 DR PROSITE; PS00626; RCC1_2; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 322 322
 FT NON_TER 322 322
 SQ SEQUENCE 322 AA; 35395 MW; DF16855EDB565ADE CRC64;

Query Match 64.9%; Score 37; DB 2; Length 322;
 Best Local Similarity 72.7%; Pred. No. 62;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAGGILHLELL 11
 ||||:||||
 Db 98 AAGGLLHSLLL 108

RESULT 15

O93P04

ID O93P04 PRELIMINARY; PRT; 322 AA.
 AC O93P04;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Chlorophyll a/b-binding light-harvesting protein Pcb
 DE (Fragment).
 GN PCB.
 OS Prochlorococcus marinus.
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
 OC Prochlorococcus.
 OX NCBI_TaxID=1219;
 RN [1]
 RC STRAIN=TAK9803-2;
 RA Garczarek L., van der Staay G.W.M., Hess W.R., Le Gall F.,
 RA Partensky F.;
 RT "Expression and phylogeny of the multiple antenna genes of the low-
 RT light adapted strain Prochlorococcus marinus SS120
 RT (Oxyphotobacteria).";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF354647; AAK69281.1; -;
 DR InterPro; IPR000932; PSIIProt.
 DR InterPro; IPR000408; Reg_chr_condens.
 DR Pfam; PF00421; PSII; 1.
 DR PROSITE; PS00626; RCC1_2; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 322 322
 FT NON_TER 322 322
 SQ SEQUENCE 322 AA; 35386 MW; BB336C0143ED3DCF CRC64;

Query Match 64.9%; Score 37; DB 2; Length 322;
 Best Local Similarity 72.7%; Pred. No. 62;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAGGILHLELL 11
 ||||:||||
 Db 98 AAGGLLHSLLL 108

Search completed: March 20, 2003, 12:34:28
 Job time : 7.37838 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:31:05 ; Search time 6.24324 Seconds
(without alignments)
256.118 Million cell updates/sec

Title: SEQID_1

Perfect score: 57

Sequence: 1 AAGGILHLELLV 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	57	100.0	12	AAE24446	Human Von Willebra
2	57	100.0	15	AAE24447	Human Von Willebra
3	57	100.0	15	AAE24448	Human Von Willebra
4	57	100.0	148	AAE24449	Human Von Willebra
5	57	100.0	203	AAE24450	Amino acid sequenc
6	57	100.0	242	AAU97641	Human aggrecanase
7	57	100.0	933	AAU79217	Human ADAM-TS-like
8	57	100.0	1120	AAU63829	Amino acid sequenc
9	57	100.0	1353	AAE24449	Human Von Willebra
10	57	100.0	1416	ABB04153	Human ADAMTS-M pol

11	57	100.0	1427	23	AAE24450	Human Von Willebra
12	45	78.9	9	23	AAE24453	Human Von Willebra
13	42	73.7	669	22	ABG11310	Novel human diagno
14	42	73.7	694	21	AAB42910	Human ORFX ORF2674
15	42	73.7	694	22	ABG11309	Novel human diagno
16	42	73.7	697	22	ABG15217	Novel human diagno
17	39	68.4	175	22	AAE24452	Human polypeptide
18	39	68.4	212	22	AAE24453	Human polypeptide
19	39	68.4	460	21	AAE24454	Human secreted pro
20	38	66.7	462	21	AAE24455	Mitochondrial proc
21	38	66.7	489	21	AAE24456	Human secreted pro
22	38	66.7	489	21	AAE24457	Human mitochondria
23	38	66.7	489	21	AAE24458	Human protein sequ
24	38	66.7	489	21	AAE24459	Arabidopsis thalia
25	37	64.9	112	21	AAE24460	Arabidopsis thalia
26	37	64.9	124	21	AAE24461	Arabidopsis thalia
27	37	64.9	139	22	ABG22850	Novel human diagno
28	37	64.9	154	21	AAE24462	Arabidopsis thalia
29	37	64.9	204	22	ABG22851	Novel human diagno
30	37	64.9	376	21	AAE24463	Arabidopsis thalia
31	37	64.9	452	21	AAE24464	Arabidopsis thalia
32	37	64.9	555	21	AAE24465	Arabidopsis thalia
33	37	64.9	732	22	AAE24466	Human protein sequ
34	37	64.9	819	22	ABG24688	Novel human diagno
35	37	64.9	819	22	ABG24689	Novel human diagno
36	37	64.9	876	20	AAE24467	Mycobacterium spec
37	37	64.9	886	20	AAE24468	Mycobacterium spec
38	37	64.9	1326	22	ABG1156	Drosophila melanog
39	36	63.2	7	23	AAE24469	Human Von Willebra
40	36	63.2	79	22	AAU44261	Propionibacterium
41	36	63.2	106	22	ABG00377	Novel human diagno
42	36	63.2	115	21	AAE24470	Human secreted pro
43	36	63.2	470	22	ABG1838	Drosophila melanog
44	36	63.2	597	22	AAE24471	Human protein sequ
45	36	63.2	665	22	AAE24472	Human colon cancer

ALIGNMENTS

RESULT 1
AAE24446
ID AAE24446 standard; peptide; 12 AA.

XX AAE24446;

DT 04-OCT-2002 (first entry)

XX Human Von Willebrand factor-cleaving protease (vWF-cp) peptide #1.

XX Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;
transgenic animal; immunisation; thromboembolic disease; pre-eclampsia;
thrombotic thrombocytopenic purpura; TTP; Henoch-Schönlein purpura;
thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
transgenic; anticoagulant.

XX Homo sapiens.

XX WO200242441-A2.

XX 30-MAY-2002.

XX 20-NOV-2001; 2001WO-EP13391.

XX 22-NOV-2000; 2000US-0721254.

XX 12-APR-2001; 2001US-0833328.

XX (BAXT) BAXTER AG.

XX Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
Schneiflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
Zimmermann K, Voelkel D;
XX

XX The invention relates to an isolated or substantially pure Von Willebrand
CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
CC solution comprising vWF with the polypeptide ligand under conditions
CC where vWF is bound to the ligand and recovering from the ligand purified
CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
CC which involves immunising an animal with vWF-cp and isolating the anti-
CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
CC producing a preparation of prophylaxis and therapy of thrombosis and
CC thromboembolic disease such as thrombotic thrombocytopenia or
CC Henoch-Schönlein purpura, preeclampsia, neonatal thrombocytopenia or
CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
CC plasmatic or recombinantly produced vWF. The invention is useful for
CC construction expression systems and generating transgenic animals which
CC express the polypeptide in vivo. The present sequence is human vWF-cp
CC N-terminal peptide.

XX SQ Sequence 15 AA;
Query Match 100.0%; Score 57; DB 23; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLV 12
Db 1 AAGGILHLELLV 12
|||||

RESULT 4
AAE24478
ID AAE24448 standard; Protein; 148 AA.
XX AAE24448;
AC AAE24448;
XX 04-OCT-2002 (first entry)
DT Human Von Willebrand factor-cleaving protease fragment #1.
DE Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;
KW transgenic animal; immunisation; thromboembolic disease; preeclampsia;
KW thrombotic thrombocytopenia; TTP; Henoch-Schönlein purpura;
KW thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
KW transgenic; anticoagulant.
XX Homo sapiens.
OS
XX WO200242441-A2.
PN 30-MAY-2002.
XX 20-NOV-2001; 2001WO-EPI3391.
XX 22-NOV-2000; 2000US-0721254.
PR 12-APR-2001; 2001US-0833328.
XX (BAXT) BAXTER AG.
PA Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
PI Scheiflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
PI Zimmermann K, Voelkel D;
XX WPI: 2002-479950/51.
DR N-PSDB; AAD39331.
XX Novel isolated or substantially purified Von Willebrand factor-cleaving
XX protease, useful for producing preparation for therapy of thrombosis
XX and thromboembolic disease such as thrombotic thrombocytopenia
XX Claim 1; Fig 3: 93pp; English.
PS The invention relates to an isolated or substantially pure Von Willebrand
XX factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
CC

CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
CC solution comprising vWF with the polypeptide ligand under conditions
CC where vWF is bound to the ligand and recovering from the ligand purified
CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
CC which involves immunising an animal with vWF-cp and isolating the anti-
CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
CC producing a preparation of prophylaxis and therapy of thrombosis and
CC thromboembolic disease such as thrombotic thrombocytopenia or
CC Henoch-Schönlein purpura, preeclampsia, neonatal thrombocytopenia or
CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
CC plasmatic or recombinantly produced vWF. The invention is useful for
CC construction expression systems and generating transgenic animals which
CC express the polypeptide in vivo. The present sequence is human vWF-cp
CC fragment.

XX SQ Sequence 148 AA;
Query Match 100.0%; Score 57; DB 23; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLV 12
Db 1 AAGGILHLELLV 12
|||||

RESULT 5
AAG63826
ID AAG63826 standard; Protein; 203 AA.
XX AAG63826;
AC AAG63826;
XX 29-OCT-2001 (first entry)
DT Amino acid sequence of a human zidnt5 polypeptide.
DE Human; zidnt5; anti-angiogenic; intestinal polypeptide; wound healing;
KW extracellular matrix interaction; tumour suppression; gamete maturation;
KW immunologic recognition; gastrointestinal irradiation; chemotherapy;
KW proteolysis; apoptosis; angiogenesis; infection; cell adhesion;
KW cell fusion; cell signalling; tumour; Crohn's disease; melanoma; trauma;
KW inflammatory bowel disease; food poisoning; degenerative disease;
KW inflammation; fertility; gamete maturation; epithelial disorder.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH Modified-site 69
FT Modified-site /note= "potential N-linked glycosylation site"
FT Modified-site 73
FT Modified-site /note= "potential N-linked glycosylation site"
FT Region 151..161
FT /note= "zinc-binding motif"
XX WO200159112-A1.
XX 16-AUG-2001.
XX 09-FEB-2001; 2001WO-US04198.
XX 10-FEB-2000; 2000US-0501806.
PR (ZYMO) ZYMOGENETICS INC.
XX Holloway JL, Sheppard PO;
PI WPI: 2001-522477/57.
DR N-PSDB; AAH74759.
XX New anti-angiogenic intestinal polypeptides, zidnt5 polypeptides, which
XX are members of disintegrin proteases, for modulating extracellular
XX matrix interaction, tumour suppression and wound healing -
XX

PS Claim 1; Page 73-74; 92pp; English.

XX The present sequence represents a human zidnt5 polypeptide. The zidnt5 polypeptide is an anti-angiogenic intestinal polypeptide. zidnt5 is used for modulating extracellular matrix interactions. zidnt5 polypeptide is useful as a tool for identifying new family members of polypeptides. zidnt5 polynucleotides are useful as probes or primers to clone 5' non-coding regions of zidnt5 gene. zidnt5 polypeptides are used for tumour suppression, gamete maturation, immunologic recognition, and growth and differentiation either working in isolation or in conjunction with other molecules in colon, small intestine, fetal lung, testis and B-cells. zidnt5 polypeptides are also useful for promoting wound healing. In the treatment of disorders associated with recovery after gastrointestinal irradiation, chemotherapy or antibody use, as anti-infectives, and extracellular matrix repair and remodeling. The polypeptides are also useful for modulating proteolysis, apoptosis, angiogenesis, infection, cell adhesion, cell fusion and signalling. The polypeptides are also useful for treating tumour formation, Crohn's disease, inflammatory bowel disease, food poisoning, melanoma, degenerative diseases, disorders related to immunity, inflammation, fertility, gamete maturation, immunology, trauma and epithelial disorders.

XX Sequence 203 AA;

Query Match 100.0%; Score 57; DB 22; Length 203;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGILHLELLV 12
Db 2 AAGGILHLELLV 13
|||||

RESULT 6

AAU97641
ID AAU97641 standard; Protein; 242 AA.

XX AC AAU97641;

XX 27-AUG-2002 (first entry)

XX Human aggrecanase protein.

XX Aggrecanase; osteopathic; antiarthritic; antiinflammatory; human; cartilage; osteoarthritis; inflammatory disease; enzyme.

XX Homo sapiens.

XX WO200233093-A2.

XX 25-APR-2002.

XX 17-OCT-2001; 2001WO-US32458.

XX 18-OCT-2000; 2000US-241469P.

XX (GEMY) GENETICS INST INC.

XX Racie LA, Twine NC, Agostino MJ, Wolfman NM, Morris EA;

XX WPI; 2002-452389/48.

XX N-PSDB; ABK52579.

XX Novel purified aggrecanase polypeptide useful for developing inhibitors and antibodies to the aggrecanase polypeptide, which are useful for treating aggrecanase-associated condition such as osteoarthritis -

XX Claim 1; Page 33-34; 41pp; English.

XX This invention relates to the cDNA and protein sequences of a novel human aggrecanase polypeptide. The protein of the invention may be used to inhibit the proteolytic activity of aggrecanase, or to inhibit

CC the aggrecanase-mediated cleavage of aggrecan in cartilage. The protein of the invention is useful for developing inhibitors of aggrecanase protein. The cDNA sequence encoding the aggrecanase protein of the invention is useful for designing probes for obtaining DNA sequences encoding other aggrecanase molecules. The cDNA sequence is also useful for detecting mRNA encoding aggrecanase in a given cell population, and thus for detecting or diagnosing genetic disorders involving the aggrecanase, or disorders involving cellular, organ or tissue disorders in which aggrecanase is irregularly transcribed or expressed. The DNA sequences may also be useful for preparing vectors for gene therapy applications. An inhibitor of the protein is useful in treating conditions characterised by degradation of articular cartilage, by blocking the enzyme's proteolytic activity. An aggrecanase protein inhibitor and a method for inhibition of its activity are useful for treating various aggrecanase-associated conditions including osteoarthritis and other inflammatory diseases. The present sequence represents the human aggrecanase protein of the invention.

XX Sequence 242 AA;

Query Match 100.0%; Score 57; DB 23; Length 242;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGILHLELLV 12
Db 42 AAGGILHLELLV 53
|||||

RESULT 7

AAU79217
ID AAU79217 standard; Protein; 933 AA.

XX AC AAU79217;

XX 15-JUL-2002 (first entry)

XX Human ADAM-TS-like protein.

XX Human; ADAM-TS-like protein; cardiovascular disorder; angina; vascular system; congestive heart failure; myocardial infarction; ischaemic heart disease; arrhythmia; hypertensive vascular disease; secondary arterial hypertension; peripheral vascular disease; embolism; chronic peripheral arterial occlusive disease; acute arterial thrombosis; inflammatory vascular disorder; chronic obstructive pulmonary disease; liver disorder.

XX Homo sapiens.

XX WO200226999-A2.

XX 04-APR-2002.

XX 26-SEP-2001; 2001WO-EP11124.

XX 28-SEP-2000; 2000US-235881P.

XX PR 25-JUL-2001; 2001US-307393P.

XX (FARB) BAYER AG.

XX Xiao Y;

XX WPI; 2002-383274/41.

XX N-PSDB; ABK49356.

XX New purified human ADAM-TS-like protein, useful for identifying modulators of protein activity for treating cardiovascular or liver disorder or chronic obstructive pulmonary disease -

XX Claim 25; Fig 2; 106pp; English.

XX The invention relates to a human ADAM-TS-like protein and the

XX Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
 PI Scheiflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
 PI Zimmermann K, Voelkel D;
 DR WPI: 2002-479950/51.
 XX Novel isolated or substantially purified Von Willebrand factor-cleaving
 PT protease, useful for producing preparation for therapy of thrombosis
 PT and thromboembolic disease such as thrombotic thrombocytopenic purpura -
 XX
 PS Claim 1; Page 64-68; 93pp; English.
 XX The invention relates to an isolated or substantially pure Von Willebrand
 CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
 CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
 CC solution comprising vWF with the polypeptide ligand under conditions
 CC where vWF is bound to the ligand and recovering from the ligand purified
 CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
 CC which involves immunising an animal with vWF-cp and isolating the anti-
 CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
 CC producing a preparation of prophylaxis and therapy of thrombosis and
 CC thromboembolic disease such as thrombotic thrombocytopenic purpura (TTP),
 CC Hensch-Schönlein purpura, preclampsia, neonatal thrombocytopenia or
 CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
 CC plasmatoc or recombinantly produced vWF. The invention is useful for
 CC construction expression systems and generating transgenic animals which
 CC express the polypeptide in vivo. The present sequence is human vWF-cp
 CC fragment.
 XX Sequence 1353 AA;
 SQ

Query Match 100.0%; Score 57; DB 23; Length 1353;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLV 12
 |||||
 DB 1 AAGGILHLELLV 12

RESULT 10
 AB04153
 ID AB04153 standard; protein; 1416 AA.
 AC AB04153;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human ADAMTS-M polypeptide.
 KW Osteoarthritis; rheumatoid arthritis; inflammatory bowel disease;
 KW Crohn's disease; asthma; Alzheimer's disease; organ transplant rejection;
 KW cachexia; allergy; cancer; leukaemia; lymphoma; osteoporosis;
 KW atherosclerosis; congestive heart failure; myocardial infarction; stroke;
 KW neurodegenerative disease; autoimmune disorder; Huntington's;
 KW Parkinson's; migraine; pain; depression; multiple sclerosis; burn;
 KW infertility; diabetic shock; gene therapy; ADAMTS-M;
 KW A Disintegrin And Metalloprotease; thrombospondin domain.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT Peptide
 FT 1..97 Location/Qualifiers
 FT /label= Prodomain
 FT /note= "The mature form of the ADAMTS-M protein is
 FT processed by furin cleavage of the prodomain"
 FT 94..97
 FT Cleavage-site
 FT 98..1416
 FT /label= Furin_cleavage_site
 FT Protein
 FT /label= Mature ADAMTS-M protein
 FT /note= "The mature form of the ADAMTS-M protein is
 FT processed by furin cleavage of the prodomain"
 FT

FT Domain 98..311
 FT /label= Metalloprotease_domain
 FT Domain 247..272
 FT /label= Zinc-binding_motif
 FT Domain 324..394
 FT /label= Disintegrin_domain
 FT Domain 410..473
 FT /label= Thrombospondin_submotif
 FT Domain 419..424
 FT /label= Heparin-binding_domain
 FT Domain 1099..1156
 FT /label= Thrombospondin_submotif
 XX
 PN EP1152055-A1.
 XX 07-NOV-2001.
 XX
 XX 24-APR-2001; 2001EP-0303706.
 PF
 XX 27-APR-2000; 2000US-200040P.
 PR
 XX (PF12) PFIZER PROD INC.
 XX
 XX Buckbinder L, Mitchell PG, Wachtmann TS, Walsh RT;
 DR WPI: 2002-084275/12.
 DR N-PSDB; ABA02549.
 XX
 PT New polynucleotide, useful in gene therapy, particularly for treating
 PT or preventing e.g. arthritis, Crohn's disease, Alzheimer's disease and
 PT organ transplant toxicity and rejection, comprises ADAMTS
 PT polynucleotide and encoded polypeptide -
 PT
 XX Claim 4; Fig 2; 31pp; English.
 PS
 XX The present sequence represents a ADAMTS protein, designated ADAMTS-M,
 CC that exhibits the characteristics of the ADAM (A Disintegrin And
 CC Metalloprotease) family of metalloproteases, and contains a
 CC thrombospondin domain (TS). The protein is encoded by the cDNA given in
 CC ABA02549. The specification describes a newly isolated polynucleotide,
 CC comprising a nucleotide sequence encoding an ADAMTS-M polypeptide as
 CC given in the specification, or a metalloprotease, disintegrin domain,
 CC prodomain or its thrombospondin submotif. The polynucleotide, polypeptide
 CC and agent are useful for manufacturing a medicament for treating a
 CC subject in need of altering activity or expression of ADAMTS-M. The
 CC polynucleotide, ADAMTS-M polypeptide and agent are useful for
 CC manufacturing a medicament for treating arthritis (osteoarthritis and
 CC rheumatoid arthritis), inflammatory bowel disease, Crohn's disease,
 CC asthma, Alzheimer's disease, organ transplant toxicity and rejection,
 CC cachexia, allergy, cancer (e.g. solid tumour cancer including colon,
 CC breast, lung, prostate, brain or haematopoietic malignancies including
 CC leukaemia and lymphoma), osteoporosis, atherosclerosis, aortic aneurysm,
 CC congestive heart failure, myocardial infarction, stroke, head trauma,
 CC spinal cord injury, neurodegenerative disease, autoimmune disorders,
 CC Huntington's disease, Parkinson's disease, migraine, pain, depression,
 CC multiple sclerosis, abnormal wound healing, burns, infertility or
 CC diabetic shock. The polynucleotide and polypeptide are also useful for
 CC diagnosing the diseases above. The polynucleotide is particularly useful
 CC in gene therapy for treating the diseases cited above.
 XX
 SQ Sequence 1416 AA;
 Query Match 100.0%; Score 57; DB 23; Length 1416;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLV 12
 |||||
 DB 98 AAGGILHLELLV 109

RESULT 11
 AAE24450

CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
CC which involves immunising an animal with vWF-cp and isolating the anti-
CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
CC producing a preparation of prophylaxis and therapy of thrombosis and
CC thromboembolic disease such as thrombotic thrombocytopenia or
CC Henoch-Schönlein purpura, preeclampsia, neonatal thrombocytopenia or
CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
CC plasmatic or recombinantly produced vWF. The invention is useful for
CC construction expression systems and generating transgenic animals which
CC express the polypeptide in vivo. The present sequence is human vWF-cp
CC N-terminal peptide.

XX
SQ Sequence 9 AA;
Query Match 78.9%; Score 45; DB 23; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGGILHLE 9
Db 1 AAGGILHLE 9
|||||

RESULT 13
ABG11310
ID ABG11310 standard; Protein; 669 AA.
XX
AC ABG11310;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #11301.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS75497.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 41669; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 669 AA;
Query Match 73.7%; Score 42; DB 22; Length 669;
Best Local Similarity 66.7%; Pred. NO. 35;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAGGILHLELLV 12
Db 10 AGGGLHLTL 21
|||||

RESULT 14
AAB42910
ID AAB42910 standard; Protein; 694 AA.
XX
AC AAB42910;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2674 polypeptide sequence SEQ ID NO:5348.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
DR N-PSDB; AAC77119.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 4526-4527; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397.

CentCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:36:13 ; Search time 19 Seconds
(without alignments)
748.837 Million cell updates/sec

Title: SEQID_1_15FUSED
Perfect score: 773
Sequence: 1 AGGGLHLELLVAVGPDVQ.....SPTWSCLITEDTGFDLGVTI 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	272	35.2	26926	1 I38344	titin, cardiac mus
2	264	34.2	15281	2 S41309	cyclosporin synthe
3	244	31.6	10797	2 T30192	probable peptide s
4	238	30.8	10223	2 T30225	polyketide synthas
5	237	30.7	9376	2 T14593	syringomycin synth
6	233	30.1	7829	2 T15789	hypothetical prote
7	229	29.6	2165	2 T21371	hypothetical prote
8	229	29.6	7463	2 T36248	CDA peptide synthe
9	225	29.1	7160	2 T27935	hypothetical prote
10	224	29.0	5255	2 T31677	bacitracin synthet
11	224	29.0	7576	2 T17428	PK506 polyketide s
12	222	28.7	5232	2 A45086	HC-toxin synthetas
13	221	28.6	6831	2 A88052	protein unc-22 (im
14	221	28.6	6839	2 S57242	twitchin (similari
15	220	28.5	6420	2 T30283	polyketide synthas
16	220	28.5	8563	2 T30226	polyketide synthas
17	219	28.3	13055	2 T16580	hypothetical prote
18	217	28.1	6805	2 S20801	titin - rabbit (fr
19	217	28.1	8243	2 T31307	type I fatty acid
20	216	27.9	4845	2 T31067	BIR repeat contain
21	214	27.7	6260	2 T30228	polyketide synthas
22	212	27.4	5035	1 I46646	ryanodine receptor
23	211	27.3	4351	2 T00252	MEGF1 protein - ra
24	211	27.3	6658	2 T13931	projectin - fruit
25	210	27.2	5032	1 A35041	ryanodine receptor
26	210	27.2	5825	2 T12117	polyprotein - fava
27	210	27.2	6486	2 T31076	tyrocidine synthet
28	209	27.0	3649	1 S18268	delta-(L-alpha-ami
29	209	27.0	4836	2 T14346	herc2 protein - mo

30	209	27.0	5147	1 IJFFTM	cadherin-related t
31	207	26.8	4613	2 T17409	polyketide synthas
32	206	26.6	4302	2 A38971	polycystic kidney
33	206	26.6	4969	2 A37113	ryanodine receptor
34	206	26.6	5037	1 A54161	ryanodine-binding
35	205	26.5	4639	1 A54794	dynenin heavy chain
36	204	26.4	5005	2 F82884	hypothetical prote
37	204	26.4	5037	2 B35041	ryanodine receptor
38	203	26.3	4967	2 S72269	ryanodine receptor
39	203	26.3	5107	2 T29144	partial CDS - Cae
40	203	26.3	5175	2 T20992	hypothetical prote
41	203	26.3	5198	2 T43290	hemiscentin precurs
42	203	26.3	5369	2 T44807	mycosubtilin synth
43	203	26.3	6359	2 T31679	bacitracin synthet
44	203	26.3	6642	2 T29757	protein UNC-89 - C
45	202	26.1	3988	1 GNWVBV	genome polypeptide

ALIGNMENTS

RESULT 1

I38344
titin, cardiac muscle [validated] - human
N:Alternate names: connectin
N:Contains: serine/threonine-specific protein kinase (EC 2.7.1.-)
C:Species: Homo sapiens (man)
C>Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 15-Sep-2000
C:Accession: I38344; I38345; S20898; S20897; S20899; S63665; S37393
R:Labelt, S.; Kolmerer, B.
S:Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Accession: I38344
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/
A:Molecule type: mRNA
A:Residues: 1-26926 <LAB1>
A:Cross-references: EMBL:X90568; NID:g1017424; PID:g1017425
R:Musco, G.; Tziatzios, C.; Schuck, P.; Pastore, A.
Biochemistry 34, 553-561, 1995
A:Title: Dissecting titin into its structural motifs: identification of an alpha-heli
A:Reference number: I38345; MUID:95119041; PMID:7819249
A:Accession: I38345
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1977-2014 <MUS>
A:Cross-references: EMBL:X83270; NID:9602579; PIDN:CAA59243.1; PID:9602580
A:Note: Conformation and properties are reported for a synthetic peptide correspondin
R:Labelt, S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A:Title: Towards a molecular understanding of titin.
A:Reference number: S20897; MUID:92258380; PMID:1582406
A:Accession: S20898
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 13597-14200, '1', 14202-14696 <LAB2>
A:Cross-references: EMBL:X64698; NID:g37192; PIDN:CAA45939.1; PID:g37193
A:Accession: S20897
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB3>
A:Cross-references: EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID:g37191
A:Accession: S20899
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 'P', 22278-22431, 'R', 22433-22448, 'O', 22450-22453, 'O', 22455-22480, 'TR', 2248
A:Cross-references: EMBL:X64697; NID:g37190; PIDN:CAA45938.1; PID:g37195
R:Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labelt, S.
J. Mol. Biol. 256, 556-563, 1996
A:Title: Genomic organization of M line titin and its tissue-specific expression in t
A:Reference number: S63665; MUID:96177761; PMID:8604138
A:Accession: S63665
A:Status: nucleic acid sequence not shown

A:Molecule type: DNA
A:Residues: 25729-26825 <KOL>
A:Cross-references: EMBL:X92412; NID:G1236761
R:Gautel, M.; Leonard, K.; Labelt, S.
EMBO J. 12, 3827-3834, 1993
A:Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentia
A:Reference number: S37393; MUID:94008990; PMID:8404852
A:Accession: S37393
A:Molecule type: mRNA
A:Residues: 26831-26926 <GAU>
R:Improt, S.; Politou, A.S.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, February 1996
A:Reference number: A66736; PDB:1TIT
A:Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
R:Fuhl, M.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A66201; PDB:1NCT
A:Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q32
A:Function:
C:Description: structural protein forming filaments in striated muscle
C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro
C:Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco
structural protein
F:24752-25008/Domain: protein kinase homology <KIN>
F:84,177,905,2276,2378,2459,2481,2563,2669,2763,2896,3088,3179,3384,3432,3628,3772,4069,
98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,13540,1
tetus predicted
F:16780,16976,17579,17602,17667,17681,17845,17899,18121,18188,18209,18336,18670,18680,18
,21900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,248
F:26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 35.2%; Score 272; DB 1; Length 26926;
Best Local Similarity 0.4%; Pred. No. 49;
Matches 101; Conservative 16; Mismatches 23; Indels 24887; Gaps 28;
Qy 9 ELL-----VAVGPDV----- 11
|||
Db 190 ELLVQGEVEPAKTKTIVSTAQISERQTRIEKIEAHFADARSATVEMVIDGAAGQOL 249
Qy 12 ----- 11
Db 250 PHKTPPRIPPKRSRPTPPSIAAKAQLARQOSPIRHSPPVRHVRAPTSPVRSVSP 309
Qy 12 -----VAVGPDV----- 18
|||
Db 310 AARISTPSIRSVRSPLLMRKTKQASTVATGPEVPPPKQEGYVASSSEAEEMRETTLTSTQ 369
Qy 19 ----- 18
Db 370 INTERWEGRYGVQEQVTISGAAGAAASVSASAAEAVATGAKEVKQDADKSAAVATV 429
Qy 19 ----- 18
Db 430 VAAVDMARVRBPVISAVEQTAQRTTTTAVHIQPAQEQVRKEAEKTAVTAVKVVVAADKAKEQ 489
Qy 19 ----- 18
Db 490 ELKSRTKEIITTKQEQMHVTHEQIRKETKTFVPPKVVISAAKAKEQETRISEITKKQKQ 549
Qy 19 ----- 18
Db 550 VTQEAIMKETRTKTVPKVIVATPKVKEQDLVSRGREGITTKREQVQITQEKMRKEAKTA 609
Qy 19 ----- 18
Db 610 LSTIAVATAKAKEQETILTRTETMATRQEQIQVTHGKVDVGKBAEAVATVVAADQARVR 669
Qy 19 ----- 18

Db 670 EPREPGLHEESYAQOTTLEYGYKERISAAKVAEPPOPPASEPHVVPKAVKRVIOAPSET 729
Qy 19 ----- 18
Db 730 HIKTDDQKMHSSQIKKTTDTLTERLVHDXKRPRTASPHFTVSKISVPKTEHGYEASIA 789
Qy 19 ----- 18
Db 790 GSAIATLQKELSATSQAQITKSVKAPTVPSETRVRAEPTPLPQFPFADTPTDYKSEAG 849
Qy 19 ----- 18
Db 850 VEYKEVGSITGTTVREERFEVLHGREAKVTETARVPAPVEIPVTPPTLVSLGNVTVI 909
Qy 19 -----FO-----AHOSDTERV- 30
|||
Db 910 EGESVTLECHISGYPSPPTVWYREDYQIESSIDFOITFOSGIARLWIREAFEDSGRFTC 969
Qy 31 ----- 30
Db 970 SAVNEAGTVSTCYLAVQVSEBFEKETTAVTEKFTTEKRFVESRDVMTDTSLTTEEQAG 1029
Qy 31 ----- 30
Db 1030 PGEPAAPYFITKVPQKLVGGSVFQGVGNPKPHVWKKSGVPLTTGTRYKYSYNKO 1089
Qy 31 ----- 30
Db 1090 TGECKLVISMTFADDAGEYTIIVRNKHGETSASALLEEADYELLMKSOQEMLYQTVTA 1149
Qy 31 ----- 30
Db 1150 FVQPEVGETAPGVFVSEVEKEYEQALIRKQAKDVTVVRTYVEDQEPHISSEERLI 1209
Qy 31 ----- 30
Db 1210 KEIYRIITKTBELLEDEGEKXAVDISSEAVESGFDLRINKYRILEGMVTFHCKMS 1269
Qy 31 ----- 30
Db 1270 GYPLPKIANYKDGKRIKHGERYOMDFLOGRASLRIPVLVPEDEGIYAFASNIKGNNAIC 1329
Qy 31 ----- 30
Db 1330 SGKLYVEPAAPLGAPTYIPTLEVPVSRISLSPRSVSRSPIRMSPARMSPARMSPARMSPA 1389
Qy 31 ----- 30
Db 1390 RMSPGRRLEETDESQERLYKPVFLKPVSKFCLEGANCDFDLKVGRPMPTETFWHDGQ 1449
Qy 31 ----- 30
Db 1450 QIVNDYTHKVIKEDGTOSLIIVPATPSDSGENTVVAQNRAGRSSISVILTVEAVEHQVK 1509
Qy 31 -----LTNLTNI----- 36
|||
Db 1510 PMFVEKLVKNVNIKEGSRLEMKVRATGNPNPDIVMLKNSDIIIVPHKYPKIRIEGTKGAAL 1569
Qy 37 ----- 36
Db 1570 KIDSTVSQDSAWYTATATANKAGRDTTTRCKNVVEVEFAEPEPERKLIIPRGTYRAKEIAAP 1629
Qy 37 ----- 36
Db 1630 ELEPLHURYGOEQWEEGDLYDKEKQKQPFKKLTSRLKRFKGAHFECLRTPISDPTMV 1689
Qy 37 ----- 36
Db 1690 VEMLHDGKPLEAANRLRMINEFGYCSLDYGVAYSRSDSGIITCRATNKYGTDHTSATLIVK 1749
Qy 37 -----GA----- 38
|||
Db 1750 DEKSLVBESQLPEGRKGLQRIEELERMAHEGALTGVTTDOKEKOKPDIVLVPBPVRVLEG 1809

QY	39	-----	38	QY	59	-----	58
Db	1810	ETARFCRVGYPPQKVNWYNGQLIRKSRFRVRYDGIHYLDIVDCSKSYDTGEVKVTAE	1869	Db	2890	HQLIIMTSTEDSAEYTFVCGNDQVSATLTVPITMITSMKLDINAEKDTITFEVTNYE	2949
QY	39	-----	38	QY	59	-----	58
Db	1870	NPEGVIEHKVLEIQOERDFRSLRRAPRPEPHVHPGKLQFEVQKVDREVDTTETKE	1929	Db	2950	GISYKWLKNGVEIKSTDKCOMRTKKLTHSLNIRNVHFGDAADYTFVAGKATSTATLYVEA	3009
QY	39	-----	42	QY	59	-----	58
Db	1930	VVKLKRABRITHEKVPESEELRSKFRTEEGYYEAITAVELSKRKXDESVEELRKT	1989	Db	3010	RHIEFRKHIDIKVLEKKRAMFECEVSEPDITVQMKDDQEOLOITDRIKIOKEYVHRL	3069
QY	43	-----	42	QY	59	-----	58
Db	1990	DELLHWTKELTSEEKKALAEBSKTIPTFKPKIELSPSMEAPKIFERIQSOTVQGS	2049	Db	3070	IPSTRMSDAGKYTVVAGGNVSTAKLFVEGRDVRIRSIKKEVQVIEKORAVPEVNEDDV	3129
QY	43	-----	45	QY	59	-----	58
Db	2050	HERVRVVGKPDCECEWYKNGVKIERSDRIYWYWPEDNVCELVRDVTAEBSASIMVKAIN	2109	Db	3130	DAHMYKDGIEINFQVQERHKYVVERIHRMPISERQSDAGEYTFVAGNRSSVTLVNA	3189
QY	46	-----	45	QY	59	-----	58
Db	2110	IAGETSSHAFLVQAKLITFTQELQDVVAKEDTMATFECETSEPFVVKVYKDGMEVH	2169	Db	3190	PEPQVQLQELQVTVQSGKPARFCAMISGRPQPKISWYKEROLLSTGFKFLHDQSYT	3249
QY	46	-----	45	QY	59	-----	58
Db	2170	EGDKYRMHSDRKVHFLSILITDTSADYSCVLVEDENVKTTAKLIVEGAVVEFVKELOD	2229	Db	3250	LLLIEAFPEDAAYTCEAKNDYGVAATTSASLSVEVPSPOEMPVYPPIITPLQDTV	3309
QY	46	-----	45	QY	59	-----	58
Db	2230	IEVPESYSGELECIVSPENIEGKWHNDVELKSNKYTITSRGRQNLTVKDVTKEDQGE	2289	Db	3310	TSEGOPARFQCRVSGTDLKVSWSKDKKIKPSRFRMTQPEDTYQLEIAEAYPEDEGTYT	3369
QY	46	-----	45	QY	59	-----	58
Db	2290	YSFVIDGKTKTKLKMPPRIATLOGLSDQKCEGDIVQLEVKVLSVESVGMKDGQEV	2349	Db	3370	FVANNAVQSVSTANLSLEAPESILHERIEQIEIEMEMKEPSSFLSABEEGLHSAEOLS	3429
QY	46	-----	45	QY	59	-----	58
Db	2350	QPSDRVHIVIDKOSHMLLIEDMTKEDAGNYSFTIPALGLSTGRVSVSDVITPLKDVN	2409	Db	3430	KINETLELLSESPVPTKFDSEKEGTGPIFIKEVSNADISMGDVATLSVTVIGIPKPKIQ	3489
QY	46	-----	45	QY	59	-----	58
Db	2410	VIEGTAVLECKVSPDVTSVKWLNDQEIKPDDRQVAIVKGTQKRLVINRTHASDEGPY	2469	Db	3490	WFFNGVLLTPSADYKVFDDGDDHSLIILFTKLEDEGEYTCMASNDYKGTICSAYLKINSK	3549
QY	46	-----	45	QY	59	-----	58
Db	2470	KLIVGRVETNCNLSEKIKIIRGLRDLTCTETQNVVFEVLSHSGIDVLWNFKDKKEIPS	2529	Db	3550	GEGHKDTETESAVAKSLEKLGCPHFLKELKPIRCAQGLPAIPEYTVVGPAPTWTWF	3609
QY	46	-----	45	QY	59	-----	58
Db	2530	SKYKIEAHGKIYKLTVLNMMKDDGKYTFYAGENMTSGKLTVAGGAIKSLPTDQTVAESQ	2589	Db	3610	KENKOLCTSVVYTIHNPNGSGTFVNDPQREDSGLYICKAENMLGESTCAELLVLLLED	3669
QY	46	-----	45	QY	59	-----	58
Db	2590	EAVFECEVANPDSKGEWLKGLPLTNIRSESDGHRRLIIAATKLLDDIGEYTYKVAT	2649	Db	3670	TDMTDTCKAKSTPEAPEDFPQTPKLGPAVEALDSEQEIATEVKDTILKAALITEENOL	3729
QY	46	-----	45	QY	59	-----	58
Db	2650	SKTSAKLVKAVKIKKTLKNTVTETODAVTVELTHPNVKGQWIKNGVVLSENEKYAI	2709	Db	3730	SYEHIAKANELSSQLPLGAQELQSILODKLTPSTREFLCINGSIHPOPLKEPSNLQL	3789
QY	46	-----	58	QY	59	-----	65
Db	2710	SVKGTIYSLRIKCAIVDESIVYGRGLGASARLHVETVKIHKPKDVTALENATVAPE	2769	Db	3790	QIVSQKTSFSGILMPPEPETOAVLSOTEKIFPSAMSIEQINSLTVEPLKTLAEPEGN	3849
QY	59	-----	58	QY	66	-----	65
Db	2770	VSVSHDTPVKWFKHSVEIKPSDKHRLVSRKVHKLMLQNI SPSDAGEYTA VVGQLECKA	2829	Db	3850	YQSSIEPPMHSVLTSAVEVLSLKEKTVSDTNREQRTVLOKQEQASALILSOSLAEGHV	3909
QY	59	-----	58	QY	66	-----	65
Db	2830	KLFVETLHITKMKNIIEVPTKTATASFECEVSHFNVPNSMWLNKNGVEIEMSEKFIIVOGKL	2889	Db	3910	ESLQSPDVMISQVNEPLVPSEHSCCTEGGKILIESANPLENAGQDSAVRIEKGKSLRFP	3969
				QY	66	-----	65

Thu Mar 20 13:50:29 2003

Db	3970	ALBEKQVLLKEHSDNVVMPDQIIESKREPVAIKKQVEQGRDLLSKESLLSGIPPEQR	4029
Qy	66	-----	65
Db	4030	LNLKIQICRALQAASVASEQGLFSEWLRNIEKVEVEANVTQEPRHIMCMLVTSKSVT	4089
Qy	66	-----	65
Db	4090	BEVTIIIEDVDPQMANLKMELRDALCAIYYEEDILTAEGRPQQAGKTSLOEEMDSFG	4149
Qy	66	-----	65
Db	4150	SQKVEPITEPEVESKYLISTEESYFNVQGRVKYLDATPVTKGASAVSDEKODESLKP	4209
Qy	66	-----	65
Db	4210	SEEKESSESSEGTAEVATVKIOEAGGLIKEDGPMIHTPLVDTVSEEGDIVHLTTSITNA	4269
Qy	66	-----	65
Db	4270	KEVNWYFENKLVPSDEKFKLQDQNTVTLVIDKNTEDHOGVEVCEALNDSGKTATSAL	4329
Qy	66	-----	69
Db	4330	TVWKRAAPVIRKRIEPIEVALGHLAKFTCEIQSAPNVRQFWKAGREIYESDKCSIRSSK	4389
Qy	70	-----	69
Db	4390	YISLEILRTQVDCGYEYTCASNEYSVCTATLTVTVPGEKKVRKLLPERKPEKEE	4449
Qy	70	-----	69
Db	4450	VVLKSLRKRPEEBEPKVEPKLEKVKPAVPEPPPKPVEEVEVPTVTKRERKIPEPTK	4509
Qy	70	-----	69
Db	4510	VPEIKPAIPLPAPEPKPEAEVNTIKPPVPEPEPTPIAAPVTVPVGVKAAKAPKEEA	4569
Qy	70	-----	69
Db	4570	AKPGPIKGVPKTPSPIEAERKLRPGSGGKPPDEAPFTYQLKAVPLKPVKEIKDIL	4629
Qy	70	-----	69
Db	4630	TESEFVGSSAIFECVLSPSTAITTMMKDGSNIRESPKHFADGKDKLHIIDVQLSDAG	4689
Qy	70	-----	69
Db	4690	EYTCVLRGNKEKTSKLVVUELPRVFKTLEBEVTVKGQPLYLSCELNKRERDVWRK	4749
Qy	70	-----	69
Db	4750	DGKIIVKXGRIVPGVIGLMRALTINDADTDAGTYTTVTENANNLECCSCVKVVEIRD	4809
Qy	70	-----	69
Db	4810	WLVKPIRDQHVKPGKTAIFACDIAKDTNPKWFKGYDEIPAEPNDKTEILRDGNHLYLKI	4869
Qy	70	-----	69
Db	4870	KNAMPEDIAEYAVEIEGKRYPAKTLGERVELLXPIEDVTIYEKESAFDAEISEADIP	4929
Qy	70	-----	69
Db	4930	GQWKLKGELLRSPPTCEIKAEGGRFLTLHKVLDQAGEVLYQALNAITAILTVKEIEL	4989
Qy	70	-----	69
Db	4990	DFAVPLKDVTPERRQARFECVLTREANVWSKGPDIKSDKFDIIADGKHHILVINDS	5049
Qy	70	-----	69
Db	5050	QPDDEGVYTAEEVGKKT SARLFTVTGIRLKFMSPLEDQTVKEGETATFVCELSHERKHVVW	5109
Qy	70	-----	69
Db	5110	FKNDAKLHTRSTVLISSBQTHKLEMKEVTLDDISQIKAQVKELSSTAOLKVLEADPYFT	5169
Qy	70	-----	69
Db	5170	VKLHDKTAVEKDEITLKCEVSKDVPVKWFKDGEIIVPSPKYSIKADGLRRLIKKADLJK	5229
Qy	70	-----	69
Db	5230	DKGEYVDCGTDKTKANVTVEARLIEVEKPLYGVFVGETAHFEIELSEPDVHGOMKLK	5289
Qy	70	-----	69
Db	5290	GOPLTASPDCEIIEBQKHILILHNCOLGTMTEVSGFOAANAKSAANLKVKEPLFITPL	5349
Qy	70	-----	69
Db	5350	SDVKVPEKDEAKPECEVSREPKTFRWLKGTQEIITGDDRFELIKDGTKHSWIKSAAFEDE	5409
Qy	70	-----	69
Db	5410	AKYMPAEDKHTSGKLIIEGIRLKLTLPLKDVTAKEKESAVFTVELSHONIRVKWPKNDQ	5469
Qy	70	-----	69
Db	5470	RLHTTRSVMOEGKTHSITFKDLSIDDTSQIRVENAGMSSSEAKLTVLSDPYFTCKLOD	5529
Qy	70	-----	69
Db	5530	YTCVEKDEVILQCEISKADAPVKWFKDGKEIKPSKNAVIKTDGKKRMLIKKALKSDIQG	5589
Qy	70	-----	69
Db	5590	YTCDCGTDKTSGLDIEDREIKLVRPLHSVEMETETARFETEISEDDIHANWKLKEAL	5649
Qy	70	-----	69
Db	5650	LQTPDCEIKEEGKIHSLVHLNCRLDQDGGVDFQAAVNVKSAHLRVKPRVIGLLRPLKDV	5709
Qy	70	-----	69
Db	5710	VTAGETATFDCELSYEDI PVWYLKGGKLEPSDKVVPSEGKVHTLTLDVKLEDADEVQ	5769
Qy	70	-----	69
Db	5770	LTAKDFKTHANLFVKEPPVEFTKPLEDQTVVEGATAVLECEVSRENKVKWFKNGTEILK	5829
Qy	70	-----	69
Db	5830	SKKYEIVADGRVRKLVIHDCPTEDIKTYTCDKDFKTSNVLNVPPHVEFLRPLTLQVR	5889
Qy	70	-----	69
Db	5890	EKEMARPECELSRENKVKWFKDGAIEIKGKYDIISKGAVRILVINKCLLDDDEAEYSCE	5949
Qy	70	-----	69
Db	5950	VRTARTSGMLTVLEEEAVFTKNLANIEVSETDTIKLVCEVSKPGAEVINYKGDDEIIEGT	6009
Qy	70	-----	69
Db	6010	RYEILTEGRKRLIVIQNAHLEDAGNYNCRLPSSRTDGVKVKVHLEAAEFISKPNLEILEG	6069
Qy	70	-----	69
Db	6070	EKAEFVCSISKESFPVQWKRDDKTLSESKYDVIADGKRLVVKDATLQDMGTVMVVG	6129
Qy	70	-----	69
Db	6130	AARAAHLTVIEKLRIVVPLKOTRVKEQQEVVFNCEVNTGAKAKWFRNEEAI FDSKYYI	6189

QY 70 ----- 69
 Db 6190 ILQKOLVYTLIRDAHLDDQANVNSLTNRGENVKSAAANLIVSEEDLRIVEPLKDIETM 6249
 QY 70 ----- 69
 Db 6250 EKKSVTFWCKVNLNVTLKWTKNGBEVPPFNDNRSYRVDKYKMLTIKDCGFPDEGEYIVT 6309
 QY 70 ----- 69
 Db 6310 AGODKSAVELLIIIEAPTEFVEHLEDQTVTEFDDAVFCQLSREKANVKWYRNGREIKEG 6369
 QY 70 ----- 69
 Db 6370 KYKFEKDSIHLIIKDCRLDDECEYACGVDRKSRARLFVEEIPVEIIRPPQDILEAPG 6429
 QY 70 ----- 69
 Db 6430 ADVVFLAELNKKVEVQWLRNNMVVQGDQKQWSEGIHRLQICDKPRDQGEYRFIAK 6489
 QY 70 ----- 69
 Db 6490 DKEARAKLELAAPKIKTADQDLVDVGKPLTMVVPYDAYPKAEAEWFKENEPLSTKTIID 6549
 QY 70 ----- 69
 Db 6550 TTAEQTSFRILEAKGDKGRYKIVLQKHGKAEGFINLKVIDVPGVPRNLEVTETFDGEV 6609
 QY 70 ----- 69
 Db 6610 SLAWEPLTDGSKIIGYVVERRDIKRKTWVLATDRAESCEFTVTGLQGGVEYLFVSA 6669
 QY 70 ----- 69
 Db 6670 RNRVGTGEVETDNPVEARSKYDVPGLNVTITDVRFGVSLTWEPPEYDGAETNYV 6729
 QY 70 ----- 69
 Db 6730 IELRDKTSIRWDAMTVRAEDLSATVTDVVEGQESFRVRAQNRIGVGKPSAATPFVKVA 6789
 QY 70 ----- 76
 Db 6790 DPTERPSPVNLTSSTQSSVOLKWEPLKDGSPILGYIIECBEGKDMIRCNKMLV 6849
 QY 77 ----- 76
 Db 6850 PELTYKVTGLEKGNKYLYRVAENKAGVSDPSEILGPLTADDAFVEPTMDLSAFKDGLEV 6909
 QY 77 ----- 76
 Db 6910 IVPNPITILVPSTGYPRPTATWCFGDKVLETGDRVKMKTLSAYAELVISPSERSDKGIYT 6969
 QY 77 ----- 76
 Db 6970 LKLENRVKTIISGEIDNVNVIARPSAPKELKFGDITKDSVHLTWEPDDGGSPLTGYVVEK 7029
 QY 77 ----- 81
 Db 7030 REVSRTWTKVMDPVTDLEFTVPDLVOGKEYLFVKYCARNKGCGEPAYVDEPNMSTPAT 7089
 QY 82 ----- 81
 Db 7090 VPDPPENVKWRDRTANSIFLTWDPKNDGGSRIKGYIVERCPGSDKWKVACGEPVAETKM 7149
 QY 82 ----- 81
 Db 7150 EVTGLEEGKWYARVKTNLROGASKPSRPTBEIQAVDQEAPEIFLDVKLLAGLTVKAGT 7209
 QY 82 ----- 81
 Db 7210 KIPLPATVTKPEPKITWTKADMILKODKRITIENVPKSTVTIVDSKRSOTGTGYIIEAV 7269

QY 82 ----- 81
 Db 7270 NVCGRATAVAVENVLDKPGPAAFDITDVTNESCLLTWNPPRDDGSKITNVVVERRATD 7329
 QY 82 ----- 81
 Db 7330 SEVMHKLSSVTKDTNFKATKLIIPNKEYIFRVAENMYGAGEPVQASPIYAKYQDFPPGPP 7389
 QY 82 ----- 81
 Db 7390 TRLEPSDITKDAVTLTWCEPDDGGSPITGYWVERLDPDTKWVRCKMPVKDITTYRVKG 7449
 QY 82 ----- 81
 Db 7450 LTNKKYRFRVLAENLAGPGKPSKSTEPILIKDIPDPWPFGKPTVKDVGKTSVRLNWK 7509
 QY 82 ----- 81
 Db 7510 PEHGGAKIESYVIEMLKTGTDEWVRVAEGVPTTQHLLPGLMEGQESYFRVRVANKAGES 7569
 QY 82 ----- 81
 Db 7570 EPSESDPVLCKEKLPPSPPPRWLEVINITKNTADLKWTVPEKDGSPITNIVVEKRDVR 7629
 QY 82 ----- 81
 Db 7630 RKGQVTDVTKTKTCTVPLTEGSLYVFRVAENAIQGSYDTEIEDSVLAKDTFTTFCP 7689
 QY 82 ----- 81
 Db 7690 PYALAVDVTKRHVDLKWEPKNDGGRPIQRYVIEKKERLGRVWVAKGTAGPCNFRVT 7749
 QY 82 ----- 81
 Db 7750 DVIEGTEVOFQVRAENEAGVHPSEPTSEILSDPTSPSPDLHLVTDAGRKHIAIAWK 7809
 QY 82 ----- 81
 Db 7810 PPEKNGSGPIIGYHVEMCPVGTEKMMRVNSRPIDKLFKEGVVDPKKEYLVRVAENAI 7869
 QY 82 ----- 81
 Db 7870 GVSEPSEISENVVAKDPCKPTIDLETHDIIIVIEGEKLSIPVPPFRAVPVPTVSHKDGKE 7929
 QY 82 ----- 81
 Db 7930 VKASDLTMKNDHISAHLEVPKSVRADAGIYITITLENKLSATASINVKVIGLPGCKDI 7989
 QY 82 ----- 81
 Db 7990 KASDITKSSCKLTWEPPEPDGTPILHYVLERREAGRRTYIPVMSGENKLSWTKDLIPN 8049
 QY 82 ----- 81
 Db 8050 GEYFRVKAANKVGGEYIELKNPVIAQDPKQPPDPVDEVHNPTAEAMTITWKPLD 8109
 QY 82 ----- 81
 Db 8110 GSKINGYIIIEKIAKEERWKRNEHLVPLTYTAKLEEGKEYOFRVRAENAAAGISEPS 8169
 QY 82 ----- 81
 Db 8170 RATPPTKAVDPIDAPKVILRTSLVKRGDEIALDASISGSPYPTITWIKDENVIVPEEK 8229
 QY 82 ----- 81
 Db 8230 KRAAPLVRRRKGEVQEEFPVLPTQRLSIDNSKKESQLRVDRSLRPHGLYMIKIVEND 8289
 QY 82 ----- 81
 Db 8290 HGIAPACTVSLDTPGPPINFVEDIRKTSVLCKWEPLDDGGSEIINVTLEKKDKTKP 8349
 QY 82 ----- 81

Db	8350	DSEWVVTSTLRHCKYSVTKLIEGKEYLFRVRAENRFGPPGVSKPLVAKDPFGPDAP	8409
QY	82	-----	81
Db	8410	DKPIVEDVTSNMLVKWNEPKDNGSPILGYWLEKREVNSTHWSRVKSNLLNALKANVDGL	8469
QY	82	-----	81
Db	8470	LEGLTYVFRVCAENAGPGKFPSPDKTAHDPISPGPPPIPRVTDTSSTTIELEMEPPA	8529
QY	82	-----	81
Db	8530	FNGGEIVGVFVKQLVGTNKNRCTERMIKVROYTVKEIREGADYKLRVSANVAGEGP	8589
QY	82	-----	81
Db	8590	PGETQPVTVAEPPQPPAVELDVSVKGGIOMAGKTLRIPAVVTCRPVPTKVWTKKEGELD	8649
QY	82	-----	81
Db	8650	KDRVVIDNVGTSKELIIKDALRKDHGRVITATNSCGSKFAAARVEVEDVPGPVLDPV	8709
QY	82	-----	81
Db	8710	VTNRKMLNWDPEDDGGSEITGFIIERKDAKMHTWRQPIETERSKCDITGLLEGOEYK	8769
QY	82	-----	81
Db	8770	PRVIANKFGCGPPVEIGPILAVDPLGPPTSPERLTYTERORSTITLDWKEPRNSGSGPI	8829
QY	82	-----	81
Db	8830	QGYIIEKRRHKDPFERVNRKLCPTTSPLENLDEHOMVEPRVKAUNEIGESEPSLPV	8889
QY	82	-----	81
Db	8890	VIQDEVPPTIKLRLSVRGDTIKVAGEPVHI PADVTGLPMPKIEMSKNETVIEKPTDAL	8949
QY	82	-----	81
Db	8950	QITKEEVSRSKTELSTIPKAVREDKGTYYVTASNRLGSVFRNVHVEYVDRSPRNLAV	9009
QY	82	-----	81
Db	9010	TDIKAESCYLTWDAFLDNGGSEITHYVIDKRDSRKAEBEVTNTAVEKRYGIWKLIPN	9069
QY	82	-----	81
Db	9070	GQYEFVRVANKYGISDECKSKVVIQDPYRLPGPPGPKVLARTKGSMLVSWTPPLDNG	9129
QY	82	-----	81
Db	9130	GSPITGYWLEKREEGSPYWSRVRAPITKVGLKGVFNVPRLLEGVKYQFRAMAINAAGI	9189
QY	82	-----CW-----	83
Db	9190	GPPSEPSDEVAGDPIFPFGPPSCPEVKDKTKSSISLGWKPKPAKDDGSP1KGYIVEMOE	9249
QY	84	-----	83
Db	9250	GTTDWKRVNEPKLITTCCEVVPNLKELRKYRFRVKAUNEAGESEPSDTTGEIPATDIOE	9309
QY	84	-----	83
Db	9310	EPEVFIDIGAQDCLVKAGSQIRIPAVIKGRPTPKSWEFDGKAMKMGVHDIPEDAQ	9369
QY	84	-----	83
Db	9370	LETAENSSVIIIECKRSHTGYSITAKNKAQCKTANCRVKVMVPGPPKDLKVSIDITRG	9429
QY	84	-----	83
Db	9430	SCRLSKWPDGDDGDR1KGYVIEKRTIDGKAWTKVNPDCGSTTFVVPDLLSEQOYFRVR	9489
QY	84	-----	83
Db	9490	AENRFIGGPPVETIORTTARDPIYPPDPPIKLIGLITKNTVHLSWKPKNDGGSPVTHY	9549
QY	84	-----	83
Db	9550	IVECLAWDTGTTKEAMROCNRDVEELOFTVEDLVEGGEYFRVKAVNAAGVSKPSATV	9609
QY	84	-----	83
Db	9610	GPCDCORPMPSPIDLKEPMEVEEGTNVNI VAKIKGVPPTLTWFKAPPKPKDNKEPVLY	9669
QY	84	-----	83
Db	9670	DTHVNKLAVDDTCTLVIQSRRSRDTGLYTTITAVNNLGTASKEMRLNLVLRPGPPVGPPIKF	9729
QY	84	-----	83
Db	9730	ESVSADQMTLSWPPPKDDGGSKITNVYIEKREANRKTWHVVSSEPKECTYTIPTKLEGHE	9789
QY	84	-----SOTIN-----	88
Db	9790	YVFR1MAQNKYGIGEPDSEPETARNLFSVPGAPDKPTVSSVTRNSMTVNWEEPEYDGG	9849
QY	89	-----	88
Db	9850	PVTGYWLEMKDTTSKRWKRVNRDPIKAMTLGVSVKVTGLIEGSDYQFRVAINAAGVGA	9909
QY	89	-----	88
Db	9910	SLPSDPATARDPIAPGPPPKVTDWTKSADLEWSPPLKDDGSKVTGYIYVEKESKEE	9969
QY	89	-----	88
Db	9970	WEXGKDEVRGTVLAVTGLKEGAFYKFRVSANVAGIGEPEVTDVIEMKDLRVSPDLQL	10029
QY	89	-----	88
Db	10030	DASVRDRIVVHAGGVIIRIAYVSGKPPPTVTNMNERTLPOEATIIETAISSSMV1KNCO	10089
QY	89	-----PEDD--- 92	
Db	10090	RSHQGVYLLAKNEAGERKKTIIVDVLDVPGVPVGTPTPLAHLNLTNESCCLTWFPSPEDDGG	10149
QY	93	-----	92
Db	10150	PITNYVIEKESDRRAWTPVYTVTRONATVOGLIOGKAYFRFAAENSIGMGPFVETSE	10209
QY	93	-----	92
Db	10210	ALVIREPITVPERPEDLEVKEVTQNTVTLTNBPKYDGGSEIINYVLESRLIGTEKPHKV	10269
QY	93	-----	92
Db	10270	TNDNLSRKYTVKGLKEGDTVEYRVSANVIGQKPSFCTKPTCKDELAPTLHLDFRD	10329
QY	93	-----	92
Db	10330	KLATIRGEAFALTCRYSGKPKPKVSWFKOBADVLEDDRTHIKTTPATLALEKIKAKRSDS	10389
QY	93	-----	92
Db	10390	GKYCVVENSTGSRKGFQCVNVVDHPGPPVGPVSFDEVTXDYMWISWKPLDDGSGKITN	10449
QY	93	-----	92
Db	10450	YIIEKVEKGVNMVPTVSASAKTTCKVSKLEGGDYIFRIHAENLYGISDPLVSDSMKAK	10509
QY	93	-----	92
Db	10510	DRFRVPADQPIVTEVTKDSALVTWNKPHDGGKPIITNYILEKRETSKRWARTKDP1H	10569

QY	93	-----	92	QY	102	-----	101
Db	10570	PYTKFRVDPDLLEGQYFRVSAENEIGIDPSPSPKPVFAKDPIAKPSPVNPPEAIDTTC	10629	Db	11650	WERVHKGSIKETHYMDRCVENQIYEFRTQNEGGESDMWKTBEVVVVKEDLQKPVLDLK	11709
QY	93	-----	92	QY	102	-----	101
Db	10630	NSVDLTWQPRHDGSKILGYVEYQKVGDEBRRANHTPESCPEYKVTGLRGQTYK	10689	Db	11710	LSGVLTKAGDTIRLEAGVRGKPPFVAVWTKDADTLTRSPRVKIDTRADSSKFSLTKA	11769
QY	93	-----	97	QY	102	-----	101
Db	10690	FRVLAVNAAGESDPAHVPEFVLVKORLEPPELIDANMAREQHIKVGDTLRLSAIIKGV	10749	Db	11770	KESDGGKYVVTATNTAGSFVAYATVNVLDKPGPVNRNLKIVDVSSDRCTVCWDPEDDGGC	11829
QY	98	-----	97	QY	102	-----	101
Db	10750	FPKVTWKEDRDAPTAKRIDVTPVSKLEIRNAAHEDGGIYSLTVENPAGSTVSKVLV	10809	Db	11830	EIONYLEKETKRMVMSTYSATVLTPTTIVTRLIEGNEYIFRVAENKIGTGPPTESKP	11889
QY	98	-----	97	QY	102	-----	101
Db	10810	LDPGPPRDLVSEIRKDSCLYLTWKEPLDDGGSVITNVYVVERRDVASAQWSPLSATSKK	10869	Db	11890	VIATKYDKPRPDPPEVTKVSKEEMTVVWNPPEYDGGKSIYGFLEKKEKHSTRWVPV	11949
QY	98	-----	97	QY	102	-----	101
Db	10870	SHFAKHLNEGQVLFRAAENQYGRPFVETPKPIKALDPLHPPGPKDLHHVDVKTEV	10929	Db	11950	KSAIPERRMKVQNLPLPOHEYOFRVKAENEIGIGBPSLPSRPVWAKDPIEPGPPPTNFRVV	12009
QY	98	-----	97	QY	102	-----	101
Db	10930	SLWNKPRDRGSGSPITGYLYVEQEBGTQDWIKFKVTNLECVVTGLQCKTYRFRVKAEN	10989	Db	12010	DTTKHSITLGMGKPYDGGAPIIGYVVMRPKIADASPDGEMKRCNAAAQLVRKEFTVTS	12069
QY	98	-----	97	QY	102	-----	101
Db	10990	IVCLGLPDITPIECOEKLVPSVELDVKLEGLVVKAGTIVRFPALIRGVPVPTAKWTT	11049	Db	12070	LDEQYEFRCVCAQNVGIGRPAELKEAIKPKLEILPEPEILDASMRKLIVIRACPIRL	12129
QY	98	-----	97	QY	102	-----	101
Db	11050	DGSEIKTDEHYTVETDNFSSVLTIKNCLRRDTGEYQITVSNAGSKTVAVHLTVLDVPGP	11109	Db	12130	FAIVRGVPAPKVTRKVGIDNVVRKGQVLDVTMAFLVIPNSTRDDSGKSLTLVNPAGE	12189
QY	98	-----	97	QY	102	-----	101
Db	11110	PTGPINILVTPHEMTISWOPPKDDGGSPVNIYVEKQTRKDTWGVSSGSKTKLIP	11169	Db	12190	KAVFVNRVLDTPGPVSDLVKSDVTKTSCHVSWAPPENDGGSQVTHYIYVEKREADRTWS	12249
QY	98	-----	97	QY	102	-----	101
Db	11170	HLOKGEYVFRVAENKIGVGPLDSTPTVAKHKFSPSPGPKPVVTDITENAAVSWTL	11229	Db	12250	TVTPEVKTSFHVTLNPGNEYFRVTAINEYGPVPTDVPKVLASDPLSEDPDRKLE	12309
QY	98	-----	97	QY	102	-----	101
Db	11230	PKSDGGSPITGYMERREVTGKWRVKNKTPIALDKFRVTGLYEGNTYEFVRFAENLAGLS	11289	Db	12310	ATEMTKNSATLAWLPLRDGGAKIDGYIISYREBEQPADRWTEYSVWKDLSLVVTGLKEG	12369
QY	98	-----	101	QY	102	-----	101
Db	11290	KPSPSDPIKACRPKPPGPPINPKLKOKSRETADLVWTKPLSDGGSPILGYVVEQKPG	11349	Db	12370	KKYKFRVAARNVAVSLPREAEGVYEAKEQLPPKILMPEQITI KAGKKLR IEAHVYKRP	12429
QY	102	-----	101	QY	102	-----	101
Db	11350	TAQWRINKDELIRQCAFVPGPLIEGNEYFRRIKAANIVGEGEPRELAESVIAKDILHPP	11409	Db	12430	HPTCKWKGDEVEVTSHLAVHKADSSILIIKOVTRKDSGYYSILTAENSSGTDQKIKV	12489
QY	102	-----	101	QY	102	-----	101
Db	11410	EVELDVTCDVITVRVQOTIRILARVKRPEPDIWTWKEGKVLVREKRVLDIQLDPRVEL	11469	Db	12490	VVMDAPGPPOPPFDISDIDADACSLSWHILPLEDGGSNITNYIYVEKCDVSRGDWVTALASV	12549
QY	102	-----	101	QY	102	-----	101
Db	11470	QIKAVRADHGKVIISAKNSSGHAQGSAINVLDPRGCPQNLKVTNVTKENCITISWENPL	11529	Db	12550	TKTSCRVGKLI PQOEYIFRVAENRFGISEPLTSPKMVAQFPFGVPSEPKNARVTKVND	12609
QY	102	-----	101	QY	102	-----	101
Db	11530	DNGSEITNFIVEYRKNQKMSIVASDVTKRLIKANLLANNEYFRVCAENKVGVGPTI	11589	Db	12610	CIFVANDRPDSGGSPITGYLIERKERNLLWVKANDTLVRSTEYPCAGLVEGLEYSFRI	12669
QY	102	-----	101	QY	102	-----	101
Db	11590	ETKTPILAINPIDRGPENLHIAKGTFFYLKWRPDPYDGGSPNLSYHVERRLKGSDD	11649	Db	12670	YALNKGSSPPSKPTTYVTARMPVDPGKPEVIDVTKSTVSLIWARPKHDGGSKIIGFV	12729
				QY	102	-----	101

```
Db 12730 EACKLPDGKWRCNTAPHQIQOEYATATGLEEKAQYOFRAIARTAVNISPPSESDPVTI 12789
Qy 102 ----- 101
Db 12790 LAENVPPRIDLSVAMKSLITVAKAGTNVCLDATVFGKPMPTVSWKKDGTLLKPAEGIKWAM 12849
Qy 102 ----- 101
Db 12850 ORNLCTLELFSVNRKSDGYTITAENSSGSKSATIKLVLDKPGPPASVKINKMYSDRAM 12909
Qy 102 ----- 101
Db 12910 LSWEPPLDGGSEITNIVDKRETSRNWAQVSATVPIITSCSVEKLIIEGHEYQFICAEN 12969
Qy 102 ----- 101
Db 12970 KYGVGDPVFTEPAIAKPNYPDPGCRDPEVISNITKDHMTVSWKPPADDDGGSPITGYLLEK 13029
Qy 102 ----- 101
Db 13030 RETQAVNMTKVNRPKPIIERTLKATLOEGTEYEFVRVTAINKAGPKPSDASKAAAYARDPQ 13089
Qy 102 ----- 101
Db 13090 YPPAPPAPPKVYDTRRSVLSWGPAYDGGSPIIIGYLVEYKRADSDNVVRNLPQNLQK 13149
Qy 102 ----- 101
Db 13150 TRFEVTGLMEDTOYQFVYAVNKIGYSDPSDVPDKHYPKILIPPEGEHADLRKTLILR 13209
Qy 102 ----- 101
Db 13210 AGVTMLRYVPVKGRPPPKITWSKPNVNLDRIGLDIKSTDFTFLRCENVNKNYDAGKYL 13269
Qy 102 ----- 101
Db 13270 TLENSCGKKEYTIVVKVLDTGPPPINVTVKEISKDSAYVWEPIIDGGSPINNVVQKR 13329
Qy 102 ----- 101
Db 13330 DAERKSWSTVTTECKTSFRVPLNEEGKSYFPRFAENEYIGDPGETRADVAKASQTPGP 13389
Qy 102 ----- 101
Db 13390 VVDLKVRSVKSSCSIGWKKPHSDGGSRIIGYVVDFTLEENKQWRVMSLSLOYSAKDLT 13449
Qy 102 ----- 101
Db 13450 EGKEYTFRVSAENENGEGTPSEITVVARDDVAPDLDLKGLPDLCLYLAKENSFRLKIPI 13509
Qy 102 ----- 101
Db 13510 KGKAPSVWKKGEDPLATDTRVSVSSAVNTLIVYDCKSDAGKYTITLKNVAGTKEG 13569
Qy 102 ----- 101
Db 13570 TISIKVKGPGIPTGPIKPFDEVTAEAMTLKMAPPKDDGGSEITNYILEKRDSVNNKWVTC 13629
Qy 102 ----- 101
Db 13630 ASAVOKTTFRVTRLHEGMEYTFRVSANKYGVGEGLKSEPIVARHPFDVPDAPPNNIVD 13689
Qy 102 ----- 101
Db 13690 VRHDSVSLTWDPKTKTGGSPITGYHLEFPKERNSLWKANKTPIRMDFKVTGLTEGLE 13749
Qy 102 ----- 101
Db 13750 EFRVMAINAGVKPSLPSEPVVALDPIDPGKPEVINITRNSVTLIWTEPKYDGHKLT 13809
Qy 102 ----- 101
```

```
Db 13810 GYIVVEKRDLPKSWMKANHVNVPECAFTVTLVEGGKYEFRIAKVNTAGIASPSESTET 13869
Qy 102 ----- 101
Db 13870 IICKOEYEAPTIIVLOPTTIKDGUTIKAGDTIVUNAISILCKPLPKSSWSKAGKDIRPSDIT 13929
Qy 102 ----- 101
Db 13930 QITSTPTSSMLTIKYATRKADAGEYITATNPFGTKVEHVKVTVLDVPGPPGPPVEISNVSA 13989
Qy 102 ----- 101
Db 13990 ERATLTWTPPLEDGGSPIKSYILEKRETSRLLTWTVVSEDIQSCRHVATKLIOGNEYIFRV 14049
Qy 102 ----- 101
Db 14050 SAVNHYGKEPQVQSEPVKMDRFGPPGPEKPEVSNVTKNATATVSHKRPVDDGGSEITGY 14109
Qy 102 ----- 101
Db 14110 HVERREKSLRMVRAIKTPVSDLRCKVTGLQEGSTYEFVSAENRAGIGPPSEASDVL 14169
Qy 102 ----- 101
Db 14170 KDAAYPPGPPSNPHVTDTTKSASLAWKPHYDGGLEITGYVVEHOKVDEAWIKDOTTGT 14229
Qy 102 -LVITRF----- 107
Db 14230 ALRITQFVVVDLOTKEKYNFRISAINDAGVGEPAVIPDVEIVEREMAPDFELDAELRRTL 14289
Qy 108 ----- 107
Db 14290 VVRAGLSIRIFVPIKGRPAPEVTWTKDNINLKNRANIENTESFTLLIIPECNRYDTGKFV 14349
Qy 108 ----- 107
Db 14350 MTIENPAGKSGFVNVRLDTARSPQLRPTDITKDSVTLHMDPLIDGSRITNYIVEK 14409
Qy 108 ----- 107
Db 14410 REATRKSYSTATTCKCHKCTYKVTGLSEGCEYFPRVMAENEYIGIETETTEPVPKASEAPS 14469
Qy 108 ----- 107
Db 14470 PPSLNMIDITKSTVSLAWPKPKHGGSKITGYVIEAQRKSGDOWTHITTVKGLECVRN 14529
Qy 108 ----- 107
Db 14530 LTEGEYTFQWMAVNSAGRSAPRESRPVIVKEOTMLPELDLRGIYOKLVIAKAGDNIKVE 14589
Qy 108 ----- 107
Db 14590 IPVLRPKPTVTKGDOILLQTORVNFETTATSTILNINECVRSDSGYPPLTARNIVGE 14649
Qy 108 ----- 107
Db 14650 VGDVITIQVHDIIPGPTGPIKFDEVSSDFVTFSDPPENDGGVPISNYVVMROTDSTTW 14709
Qy 108 ----- 107
Db 14710 VELATTVIRTTYKATRLTTGLEQYQFRVKAQNRVYGVGPGITSAMIVANYPFKVPGPCTPQ 14769
Qy 108 ----- 107
Db 14770 VTAVTKDSMTISWHEPLSDGSPILGVHVERKERNGLWQTSKALVPGNIFKSSGLTDG 14829
Qy 108 ----- 107
Db 14830 IAYEFRVIAENMAGSKSPSEPMLALDPIDPGKPVPLNITRHTVTVLKWAKEYTGGF 14889
Qy 108 ----- 107
Db 14890 KITSYIVVEKRDLPNGRWLKNFNSILENFTVSLGTEDAAVEFRVIAKNAAGAI5PSPSEP 14949
```

Qy 108 ----- 107
Db 14950 SDAITCRDDVEAPKIKVDYKFDTVILKAGEAFRLEADVSRPPTMEWSKDGELEGT 15009
Qy 108 ----- 107
Db 15010 KLEIKIADFSTNLNKNKSTRDSGAYTLTATNPGGFAKHFNVKVLDRPGPEGLAVTE 15069
Qy 108 ----- 107
Db 15070 VTSEKVLSPFPPLDDGGAKIDHIYIVQKRETSRLAWTNVASEVQVTKLVTKLLKGNEYI 15129
Qy 108 ----- 107
Db 15130 FRMAVNKYGVGEPELESEPLAVNPGPPDKNPEVTITKDSMVVCHGHPDSGGSEI 15189
Qy 108 ----- 107
Db 15190 INVIVERRDKAGORWIKCNKKTLDLRYKVSGLTEGHEYEFRIMAENAAGISAPSPSPF 15249
Qy 108 ----- 107
Db 15250 YKACDVFVKPGPPGNRVLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEDEWQIVT 15309
Qy 108 ----- 107
Db 15310 PPAGLKATSYTITGLTENQEYKIRIYAMNSEGLGEPALVPGTPKAEDRMLPPEIELDADL 15369
Qy 108 ----- 107
Db 15370 RKVWTIRACCTLRLFPVPIKGRDPPEVKWARDHGESLDKASIESAGSYTLIIVGNVNRFS 15429
Qy 108 ----- 107
Db 15430 GXYILTVNSSGSAFVNVNVLDTGPPQDLKVKVEVTKTSVTLTWDPPLDGGSKIKNY 15489
Qy 108 ----- 107
Db 15490 IVEKRESTRKAYSTVATNCHTKTSWKVDQLOEGCSYFRVLAENEYIGLPAETAESVKAS 15549
Qy 108 ----- 107
Db 15550 ERPLPGKITLMDVTRNSVSLWEKPEHDGGRILGYIVEMOTKSGDKWATCATVKVTEA 15609
Qy 108 ----- 107
Db 15610 TITGLIOGEEYFRVSAQNEKGISDPRQLSVPIAKDLVIPPAFKLLFNTFTVLAGEDLK 15669
Qy 108 ----- 107
Db 15670 VDVPPFIGRPTPAVTHKDNVPLKQTTRVNAESTENNSLLTIKDAREDVGHVVKLTNSA 15729
Qy 108 ----- 107
Db 15730 GEAIETNLVILDKPGPTGPVKWDEVADTSITLSWGPVKPYDGGSSINNYIVEKRDSTT 15789
Qy 108 ----- 107
Db 15790 TWOIVSATVARTTIKACRLKTGCEYQFRIAAENRYGKSTYLNSEPTVAQYPPKVPQPGT 15849
Qy 108 ----- 107
Db 15850 PVTLSRDSMEVQWNEPISDGGSRVIGYHLERKERNILWVKLNTPIPTQTKFKTTGLE 15909
Qy 108 ----- 107
Db 15910 EGVEYFRVSAENIVGIGKPSKSECYVARDPCDPPRPEALIVTRNSVTLQWKPTYDG 15969
Qy 108 ----- 107
Db 15970 GSKITGYVEKKELPEGRWMAKSFNTIIDTHFEVTGLVEDHRYEFRVIARNAAGVFSFS 16029

Qy 108 ----- 107
Db 16030 ESTGAITARDEVDPRISMDFPKYKMDTIVHAGESFKVDADIYGPPIPTIOWIKGDQLSN 16089
Qy 108 ----- 107
Db 16090 TARLEIKSTDPAFSLSVKDAVRVDSGNVILKAKNVAGERSVTNVKVLDRPGPEGPVVI 16149
Qy 108 ----- 107
Db 16150 SGVTAECTLAWKPLQDDGSDIINYIVERRETSRLVMTVVVDANVQTLSCVTKLLEGNE 16209
Qy 108 ----- 107
Db 16210 YTFRIMAVNKYGVGEPELESEPVVAKNPVVPDAPKAEVTTVTKDSMIVMWERPASOGGS 16269
Qy 108 ----- 107
Db 16270 EILGVULEKROKEGIRWTRCHKRILGELRLAVTGLIENHDYEFVRVSAENAGLSEPPSPS 16329
Qy 108 ----- 107
Db 16330 AYQACDPIYKPGPPNPKVIDITRSSVFLSWSKPIYDGGCEIOGYIVEKCDVNVGWTM 16389
Qy 108 -----DL 109
Db 16390 CTPPTGINKTNIIEVKLEKHEYNFRICAINKAGVGEHADVPGPPIIVEEKLEAPDIDL 16449
Qy 110 EL----- 111
Db 16450 ELRKIINRAGSURLFPVPIKGRPTPEVKWKVDGEIRDAIIVTSSFTSLVLDNVNRY 16509
Qy 112 ----- 111
Db 16510 DSGKYTLTLENSGTSKSAFVTVRVLDTSPVPLKVTBITKDSVSIWEPPLDGGSKIK 16569
Qy 112 ----- 111
Db 16570 NYIVEKREATRKYAAVVTNCHKNWKIDLOEGCSYFRVTAENEYIGILPAQTADPIK 16629
Qy 112 ----- 111
Db 16630 VAEVPPPGKITVDVTRNSVSLWTKPEHDGSKIIOYIVEMQAKHSEKSEKARVKS 16689
Qy 112 ----- 111
Db 16690 QAVITNLTOGBEYLFRVVAVNNEKGRSDPRSLAVPIAKDLVIEPDVKPAPFSYSVQVGD 16749
Qy 112 ----- 111
Db 16750 LKMEVPISGRPKPTITWTKDGLPLKQTTIRINVTSLDLTTLISIKETHKDDGGYGITVAN 16809
Qy 112 ----- 111
Db 16810 VVGQXTASIEIVTLDKPDPKPGVKFDVDSAESITLSWNPPLYTGCCQITNVIQKRDIT 16869
Qy 112 ----- 111
Db 16870 TTVMDVSATVARTTLKVTCLKTGTGYQFRIFAENRYGQSFALSDPIVAQYPPKEPQP 16929
Qy 112 ----- 111
Db 16930 GTPFATAISKDSMVJQWHEPVNNGSRVIGYHLERKERNILWTKVNTIHDTOFKAON 16989
Qy 112 ----- 111
Db 16990 LEEGIEYFRVVAENIVGVKASKNSECVVARDPCDPPGTPEPIMVKNRNEITLQWTKPVY 17049
Qy 112 ----- 111
Db 17050 DGSMTIGYVEKROLDPGRWMAKSFNTNVIETQFTVSLGLEDQRYEFRVIARNAAGAIK 17109
Qy 112 ----- 111

[illegible]

Qy 134 ----- 133
Db 19330 GIEIERARTEIISTDNHTLLTVKDCIRRDGTQOYVLTAKNAGTRSAVNCVKLDKPGPP 19389
Qy 134 ----- 133
Db 19390 AGPLEINGLTAEKCSLSWGRPOEDGGADIDYYHRKKRETSHLAWTICEGELQMTSCVKTK 19449
Qy 134 ----- 133
Db 19450 LUKGNEYIFRVTVGNKYGVGEPLSAIKALDPFTVPSPPTSLEITSVTKESMTLCWSRP 19509
Qy 134 ----- 133
Db 19510 ESDGSEISGYIIERREKNSLRWRJNKKPVYDLRVKSTGLREGCEYEYRVYAENAGLS 19569
Qy 134 ----- 133
Db 19570 LPSETSLIRAEDPVFLPSPSPKPIVDGKTTITIAWKPLFDGAPITGTVVEYKSD 19629
Qy 134 ----- 133
Db 19630 DTDWKTISQSLRGTEYTIISGLTTGAEYFRVKSVMKVGASDPSSDPQIAKEREPEPLF 19689
Qy 134 ----- 133
Db 19690 DIDSEMRKTLIVKAGASFTWTVFRRGPVNVLSKPDPTDLRTRAVVDTTDSRTSLTIEN 19749
Qy 134 ----- 133
Db 19750 ANRDSKGYTLTIQNLVLSAASLTLLVKVLDTPGPTNITVQDVTKESAVLSMDVPENDGG 19809
Qy 134 ----- 133
Db 19810 APVKNVHIEKREASKAWSVTNNCNRLSYKVTNLOEGAIYYFRVSGENEFVGVPATK 19869
Qy 134 ----- 133
Db 19870 EGVKITEKPPPEKJGVTISKDSVSLTWLKPHEHGGSRIVHYVVEALEKQKNWVCAV 19929
Qy 134 ----- 133
Db 19930 AKSTHHVVSGLRESEYFFRVFAENQAGLSDPRELLLPVLIKEQLEPEIDMKNFPSHTV 19989
Qy 134 ----- 133
Db 19990 YVRAGSNLKVDIPISGKPLKVTLSDRGVPLKATMRFNTEITAENLTINLKESVTADAGR 20049
Qy 134 ----- 133
Db 20050 YEITAANSSGTTKAFINIVVLDROPPTGPVVISDITEESVTLKNEPPKYDGSQVTNYI 20109
Qy 134 ----- 133
Db 20110 LLKRETSAVTVESATVARTMKVMKLTGTGEEYQPRIKAENRFGISDHIDSACVTVKLP 20169
Qy 134 ----- 133
Db 20170 YTPGPSTPMVNTVRESITVWHEPVNSGSAVVGHYHLEMKDRNSILWQKANKLVIRT 20229
Qy 134 ----- 133
Db 20230 THFKVTTISAGLIYEFRYAENAGVKPSHPSEPLAIDACEPPNRNVRTIDISKNSVL 20289
Qy 134 ----- 133
Db 20290 SWQOPAFDGGSKITGYIVERDLDPGRWTKASFNTVTTQTTISGLTQNSQYERFVPARN 20349
Qy 134 ----- 133
Db 20350 AVGSISNPSEVVGPICTIDSYGGPVIDLPLEYTEWVKYRAGTSVKLRAGISGKPAPTIEW 20409

Qy 134 ----- 133
Db 20410 YKDDKELQTNALVCVENTTDLASILKADRLNSGCYELKLRNAMASASATIRVQILDKP 20469
Qy 134 ----- CLIT 137
Db 20470 GPPGPIBFKTVTAEBKITLLWRPPADGGAKITHYIVEKRETSRVVMSVSEHLEECIIT 20529
Qy 138 ----- 137
Db 20530 TTKIIKNEYIFRVRAVNNKYGIGEPLESDSVVAKNAFVTPGPGIPEVTKITKNSMTVM 20589
Qy 138 ----- 137
Db 20590 SRPIADGGDISGYFLEKRXKSLGWFKVKETIRDTROKVTGLTENSQYQYRVCAVNA 20649
Qy 138 ----- 137
Db 20650 GQGFSEPFSEFYKAADPIDPPGPPAKIRIADSTKSSITLGMSPKPVYDGGSAVTGVVEIR 20709
Qy 138 ----- 137
Db 20710 QGEEBEWTVSTKGEVRTTEYVVSNLKPGVNYFRVSAVNCAGOGGEPTEMNEPVQAKDIL 20769
Qy 138 ----- 137
Db 20770 EAPEIDLVALRTSVIAKAGEDVQVLIPFKGRPPPTVTRKDEKMLGSDARYSIENTDSS 20829
Qy 138 ----- 137
Db 20830 SLLTIPQVTRNDTGKYLITENGVEPKSVSVKVLDTPAACQKLVKHVSRGTVTLLW 20889
Qy 138 ----- 137
Db 20890 DPPLIDGSPINYYIEKDATKRTWSVSHKCSSTSFKLIDLSEKTPFFFRVLAENEIG 20949
Qy 138 ----- 137
Db 20950 IGECETTEPVKAAEVPAPIRDLNMDSTKTSVILSWTKPDPDGGSVITEYVVERKGE 21009
Qy 138 ----- 137
Db 21010 QTWSHAGISKTEIEVSOLKEQSVLEFRVFAKNEKGLSDPVTIGITVKELIITBEVDLS 21069
Qy 138 ----- 137
Db 21070 DIPGAQVTVRIGHNVHLELPYKGPSPISMLKDGPLKESEFVRFKSTENKITLSIKVA 21129
Qy 138 ----- 137
Db 21130 KKEHGGKVTVILDNAVCRIVPITVITLGPSPKGPPIRFBDEIKADSIVLSWDVPEDNGG 21189
Qy 138 ----- 137
Db 21190 GEITCYSIEKRETSOTNNWMCSSVARTTFKVPNLVKOAEYOFVRAENRYGVSQPLVSS 21249
Qy 138 ----- 137
Db 21250 IIVAKHQFRI PGPBGKPVINYNTSDGMSLTWDA PVYDGGSEVTGFHVEKKERNLSILWQKV 21309
Qy 138 ----- 137
Db 21310 NTSPISGREYRATGLVEGLDYQFRVYAENSAGLSSPSPDKFTLAVSPVDPPTPDYIDV 21369
Qy 138 ----- 137
Db 21370 TRETTIKWNPPLRDGSKIVGYSIEKROGNERWVRNFTDVSECOYTVTGLSPGDRYEF 21429
Qy 138 ----- 137
Db 21430 RIARNAVGTISPPOSSGIIIMTRDENVPPIVEFGPEYFDGLIIKSGESIRIKALVQGRP 21489
Qy 138 ----- 137

Db	21490	VPRVTWFKDGVIEKRMMEITNVLGSTSLFVRDATROHRCVYTVVEAKNAGSAGAEIKV	21549
Qy	138	-----	137
Db	21550	KVQDTPCKVVGPIRFTNITGEKMTLMDAPLNDGCAPITHYIEKRETSRLAWALIEDKC	21609
Qy	138	-----	137
Db	21610	EAOSYTAIKLINGNEYQFRVSANVKFGVGRPLDSPVVAQIQYTVDPAGIPEPSNITGN	21669
Qy	138	-----	137
Db	21670	SITLTWARPESDGGSEIQOYLIERREKKSTRWVKVISKRPISETRPKVTGLTEGNEYEPH	21729
Qy	138	-----	137
Db	21730	VMAENAAVGVPASGISRLIKCREPNPPPTVVVKVTDTSKTTVSLEWSKPVFDGMEII	21789
Qy	138	-----	137
Db	21790	GYIEMCKTDLGDMHKVNAEACVTRYTVTDLQAGEYKFRVSAINGAGKGDSCVETGTI	21849
Qy	138	-----	137
Db	21850	KAVDRLTAPELDIDANFKQTHVRAGASIRLFIAYQGRPTPTAVMSKPDNSLSURADIHT	21909
Qy	138	-----	137
Db	21910	TDSFSTLTVCNCRNDAGKTYLTVENNSGSKSIITFTVKVLDTPGPPGPIITPKDVTGRSAT	21969
Qy	138	-----	137
Db	21970	LMWDAPLDGGARIIHHVYVVEKREASRRSQVISEKTRQIIFKVNDLAEGVPYFRVSAVN	22029
Qy	138	-----	137
Db	22030	EYGVGEPEMPEPIVATEQPAPRRLDVVDTSKSSAVLAWLKPDDHGGSRITGYLLEMQR	22089
Qy	138	-----	137
Db	22090	KGSDLWEAGHTKQLTFTVERLVEKTEYFRVKAKNADAGYSEPREAFSSVIIKEPQIEPT	22149
Qy	138	-----	137
Db	22150	ADLTGITNOLITCKAGSPFTIDVPISGRPAKVWTKLEEMRLKETDRVSIITTKDRTTLT	22209
Qy	138	-----	137
Db	22210	VKDSMRGDSGRYFLTLENTAGVKTFSTVWVIGRPGVTPGPIEVSSVSAESCVLWSGBPK	22269
Qy	138	-----	137
Db	22270	DGGGTEITNYIVEKRESGTTAWQLNVSSVKRTOIKVTHLTKYMEYSFRVSSENRFGVSKP	22329
Qy	138	-----	137
Db	22330	LESAPIIAEHPPVPSAPTRPREVYHVSANAMSIIRWEEPHYDGGSKIIGYWVEKKERNITL	22389
Qy	138	-----	137
Db	22390	WYKENKVPCLCNYKVTVGLVEGLEQFRTYALNAAGVSKASEASRPINAQNPVDAPGRPE	22449
Qy	138	-----	137
Db	22450	VTDVTRSTVLSIWSAPAYDGSKWGYIIERKPVSEVGDRWLKNCYTIIVSDNFTVTAL	22509
Qy	138	-----	137
Db	22510	SEGDTYEFRLAKNAAGVTSKSESTGPVTCRDEYAPPAELDARLHGLDVTIRAGSDLV	22569
Qy	138	-----	137
Db	22570	LDAAVGGKPEPKIINTKGDKELDCEKVSLOYTGKRATAVIAIKFCDRSDSGKYTLTVKNAS	22629
Qy	138	-----	137
Db	22630	GTKAVSMVKVLDSPGCGKLTVSRVTQEKCTLAWSLPOEDGGAETHYIVERRETSRLN	22689
Qy	138	-----	137
Db	22690	WVIVEGECPTLSVYVTRLIKNNEYIFRVRVANKYGPVPESEPIVARNSTIPSPGP	22749
Qy	138	-----	137
Db	22750	EEVGTKEHIIIIQWTKPESDGGNEISNYLVDKREKSLRWTRVKNKYVYVDTRLKVTSLM	22809
Qy	138	-----	137
Db	22810	EGCDYQFRVTAVNAAGNSBPSERSNFISCREPSYTPGPPSAPRVVDTTKHSISLAWTKPM	22869
Qy	138	-----	137
Db	22870	YDGGTDIVGVLEMOEKTDQWYRVHTNATIRNTEFTVPDLKMGQKYSFRVAANVKGMS	22929
Qy	138	-----	137
Db	22930	EYSESAEIEPVERIEIPOLELADLLKKTVTIRAGASLRLMVSVSGRPPVITWSKQID	22989
Qy	138	-----	137
Db	22990	LASRAIIDTTESYSLIIVDKVNRDAGKVTIEAENOSGKKSATVLVKVYDTPGCPSPVKV	23049
Qy	138	-----	137
Db	23050	KEVSRDSVTITWBIPTDGGAPINNYIVEKREAAAFKTVTTKCSKTLVRIISGLVEGTM	23109
Qy	138	-----	137
Db	23110	HYFRVLPENIYGEGECETSDAVLVSEVPLVPAKLEVDVTKSTVTLAWEKPLYDGSRL	23169
Qy	138	-----	137
Db	23170	TGYVLEACKAGTERMKVVTLKPTVLEHTVTSLEGEQVLFIRAQNEKGVSEPRETVTA	23229
Qy	138	-----	137
Db	23230	VTQDLRVLPTIDLSTMPQKTIHVPAGRPVELVIPIAGRPPPAASWFFAGSKLRESERV	23289
Qy	138	-----EDTG-----	141
Db	23290	WETHKVAKLTIRETITRTDGEYTLKXVTGTTSETIKVIIIDKPGPTGPIKIDEIDA	23349
Qy	142	-----	141
Db	23350	TSITISWEPPELDGGAPLSGVVVEQDAHRPGWLPVSESVTRSTFKFTRLTEGNEYVFRV	23409
Qy	142	-----	141
Db	23410	AATNRFSGSYLOSEVIECRSSIRIPGPETLOIFDVSRDGMTLTWYPEDDGGSQVTGY	23469
Qy	142	-----	141
Db	23470	IVERKEVRADRVRVKNKVPVMTYRSTGLTEGLEYEHRVTAINARGSKPSRPSKPIVA	23529
Qy	142	-----	141
Db	23530	MDPIAPGKPNQPRVTDTRTSVSLAWSVPEDEGSKVTGYIEMQVQDOHEMTKCTTP	23589
Qy	142	-----	141
Db	23590	TKIREYTLHLPOGAEYRFRVLACNAGGPGEPAEVPQTKVTEMLEYPDYELDERYQEG	23649
Qy	142	-----	141
Db	23650	FVROGGVIRLTIPIKGRPFPICKWKTEGQDISKRAMIATSETHTELVIKEADRGDSCTYD	23709

```
QY 142 ----- 141
Db 23710 LVLENCKGKAVYKVRVIGSPNSPEGPLEYDDIQVRVSVRPPADDDGADILGYILE 23769
QY 142 ----- 141
Db 23770 RREVPKAAWYIDSRVRGTSLVVKGLKENVEYHFRVSAENQFGISKPLKSEEPVTPKTEL 23829
QY 142 ----- 141
Db 23830 NPPEPSNPPEVLDTKSSVLSWSRPKDDGSRVTGYIYIERKETSTDKVVRHNKTOITT 23889
QY 142 ----- 141
Db 23890 TMYTVGLVPDAEYQFRITANDVGLSETSPASEPVVCKDPFDPKPSQPGELEILSISKDS 23949
QY 142 ----- 141
Db 23950 VTLQWEKPECDGCKEILGYWVEYRQSDSAWKSNKERIKDKQFTIGGLLEATEYEFVRV 24009
QY 142 ----- 141
Db 24010 AENETGLSRPRRTAMSIKTKLTSGEAPGIRKEMKDVTTKLGEAAQLSCQIVGRPLPDIKW 24069
QY 142 ----- 141
Db 24070 YRFGKELIQSRKYKMSDGRHTLTVWTERQEDGVYTCIATNEVGEVETSSKLLLOATP 24129
QY 142 ----- 141
Db 24130 QHPGYPLKEKYYGAVGSTLRLHWYIGRPVPMPTWPHGQKLLQNSENIENTEHVTHL 24189
QY 142 ----- 141
Db 24190 VMKNVORKTHAGKYKVQLSNVFGTVDAILDVEIQDKPKPTGPIVIEALLKNSAVISWKP 24249
QY 142 ----- 141
Db 24250 PADGGSWITNYVVEKCEAKEGAEWQLVSSAISVTTCRIVNLNENAGYIFRVSAQNTFGI 24309
QY 142 ----- 141
Db 24310 SDPLEVSSVVIKSPPEKPGAPKPTITAVTKDSCVVMKPPASDGGAKIRNVYLEKREK 24369
QY 142 ----- 141
Db 24370 KONKWIWVTEIETVFSVKNLIEGLEYEFVRVKCENLGESEWSEISEPITPKSDVPIQ 24429
QY 142 ----- 141
Db 24430 APHPKEELNLRVYQSNATLVCKVCHPKPIVKNYRQGEIADGLKYRIQEFKGYHQ 24489
QY 142 ----- 141
Db 24490 LIASVTDDATVYQVRATNGGVSQSGTASLEVEVPKIHLPKTLQMGAVHALRGEWS 24549
QY 142 ----- 141
Db 24550 IKIPFSGKPDVITWQKQDLIDNNGHYQVITRSFTSLVPFNGVERKDAGFYVVCANR 24609
QY 142 ----- 141
Db 24610 FGIDQKTVELDVADVPDPGRGVKVSASRDSVNLTWTEPDSGSKITNIVEKATTAE 24669
QY 142 ----- 141
Db 24670 RWRVQARETRYVINLFGKTSYQFRVIAENKFGLSKPSPEPTITKEDKTRAMNDE 24729
QY 142 ----- 141
Db 24730 EVDETREVSMTKASHSSTKELYEKYKYMIAEDLGRGEFGIVHRCVETSSKKTVMKFKVK 24789
```

```
QY 142 ----- 141
Db 24790 TDQVLVKEISILNIARHNILHLHESFESMEELVMIFEFGSLDIFRINTSAFELNER 24849
QY 142 ----- 144
Db 24850 EIVSVHOVCEALQLHSHNIGHFDIRPENIYYQTRRSSTIKIIEFGQARQLKPCDNFRL 24909
QY 145 ----- 144
Db 24910 LFTAPEYVAPEVHQHDVVSTATDMWSLGLTVVLLSGINPFLAETNQOIENIMNAEYTF 24969
QY 145 ----- 144
Db 24970 DEEAFKEISIEAMDFVDRLLVKERKSRMTASEALQHPMLKOKIERVSTKVIRTLKHRRYY 25029
QY 145 ----- 144
Db 25030 HTLIKKDLNMVVSAAIRISCGGAIRSQGVSVAKVVASIEIGPVSQIMHVAEGEGHVK 25089
QY 145 ----- 144
Db 25090 YVCKIENVYDQSTQVTWYFGRQLENSEKYEITYEDGVAIYVKDITKLDGTYRCKVND 25149
QY 145 ----- 144
Db 25150 YGEDSSYAELFVKGVREVVYCYCRRTMKIKRRTDMRLERPEFTLPLYNKTAIVVGEN 25209
QY 145 ---GVTI 148
Db 25210 VRFGVTI 25216

RESULT 2
S41309
Cyclosporin synthetase - cyclosporin fungus
C:Species: Tolypocladium inflatum (cyclosporin fungus)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 03-Nov-2000
C:Accession: S45487; S41309
R:Weber, G.; Schoergerdorfer, K.; Schneider-Scherzer, E.; Leitner, E.
Curr. Genet. 26, 120-125, 1994
A:Title: The peptide synthetase catalyzing cyclosporine production in Tolypocladium n
A:Reference number: S45487; MUID:95094306; PMID:8001164
A:Accession: S45487
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-15281 <WEB>
A:Cross-references: EMBL:Z28383; NID:G440168; PIDN:CAA82227.1; PID:G440169
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993
A:Note: only a part of the translation is shown
A:Note: the source is designated as Tolypocladium inflatum
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: cyclosporin fungus cyclosporin synthetase; acetate-CoA ligase homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:532-1008/Domain: acetate-CoA ligase homology <ACLI>
F:1029-1096/Domain: acyl carrier protein homology <ACP1>
F:1618-2069/Domain: acetate-CoA ligase homology <ACLI2>
F:2527-2594/Domain: acyl carrier protein homology <ACP2>
F:3115-3553/Domain: acetate-CoA ligase homology <ACLI3>
F:4014-4081/Domain: acyl carrier protein homology <ACP3>
F:4601-5057/Domain: acetate-CoA ligase homology <ACLI4>
F:5506-5573/Domain: acyl carrier protein homology <ACP4>
F:6094-6546/Domain: acetate-CoA ligase homology <ACLI5>
F:7003-7070/Domain: acyl carrier protein homology <ACP5>
F:7591-8042/Domain: acetate-CoA ligase homology <ACLI6>
F:8063-8130/Domain: acyl carrier protein homology <ACP6>
F:8652-9157/Domain: acetate-CoA ligase homology <ACLI7>
F:9558-9625/Domain: acyl carrier protein homology <ACP7>
F:10146-10586/Domain: acetate-CoA ligase homology <ACLI8>
F:11055-11122/Domain: acyl carrier protein homology <ACP8>
F:11635-12106/Domain: acetate-CoA ligase homology <ACLI9>
F:12127-12194/Domain: acyl carrier protein homology <ACP9>
```

F:12715-13159/Domain: acetate-CoA ligase homology <AL10>
F:13623-13690/Domain: acyl carrier protein homology <AC10>
F:14213-14676/Domain: acetate-CoA ligase homology <AL11>
F:14698-14765/Domain: acyl carrier protein homology <AC11>
F:1060,2558,4045,5537,7034,8094,9589,11086,12158,13654,14729/Binding site: phosphopanted

Query Match 34.28; Score 264; DB 2; Length 15281;

Best Local Similarity 0.84; Pred. No. 17;

Matches 102; Conservative 24; Mismatches 21; Indels 12475; Gaps 32;

Qy 1 AAGGIL----- 6
||| :|
Db 2098 AAGHVEIGTGTGMVLFNLGQAGLSYIGLEPSQSAQVFNKAAQTPGLEKQAQVHGT 2157
Qy 7 ----- 6
Db 2158 AMDTGRSALSPDLIVNSVAQYFSPREYLAEVVEALVRIPGVRRIRFGDMRTYATHKDF 2217
Qy 7 -----HLELL----- 11
||| :|
Db 2218 LVARAVHTNGSKVTRSKVQEQVEARLEELLEVDPAPFTSLKESLEEIEHVEILPKNM 2277
Qy 12 ----- 11
Db 2278 KVNNELSSRYGAVLHIRNHNQNSRSHKINAESWIDFASSQMDRQGLARLLKENKDAE 2337
Qy 12 ----- 11
Db 2338 SIAVENIPYSKTIVERHIAKSLADHDGDDTHSSIDGVAMISAAREKASQCPSLDVHDLV 2397
Qy 12 ----- 11
Db 2398 QLAEDAGPRVEVSWARQSQNGALDVFFHHFOPTENESRALVDPTDYKGOQARSLTNRP 2457
Qy 12 ----- 11
Db 2458 LQVESRRIEAQVREQLQVLLPAYMPIPARIVVLQNMPLNTSGKVDKRELTLRAKVTAART 2517
Qy 12 ----- 11
Db 2518 PSSELVAPRDSIEAIIKCEKFDVLGVVEGITDNFFNVGHSLLATKLAARLSRQLNAQIA 2577
Qy 12 ----- 11
Db 2578 VKDIFDRPVIADLAATIOQDTTEHPILPTSYTGPVEQSFAGRLWFLDQLNVGATWYLM 2637
Qy 12 ----- 11
Db 2638 PFAVRLGPLVVSALAAALLAEERHETLRTTFIEQEGIGQVHPFAPKELRVIDVSGE 2697
Qy 12 ----- 11
Db 2698 EESTIQILEKEQTFFNLASEPGFRLALLKTGEDEHILSTVMHHAISDGWSVDIFQOEI 2757
Qy 12 ----- 11
Db 2758 GOFYSAILRGHDPQLAIPLSIQYRDFATWQRQIFQVAEHRRLQAYWKQLADNKPALL 2817
Qy 12 ----- 11
Db 2818 TDFKRPMLSGRAEIPVVVDGLIYEKLQDFCRIRQVTAFTVLLAAFRAAHYRMTGTEDA 2877
Qy 12 ----- 11
Db 2878 TIGTPIANNRPELEGLIGFFVNTQCMRITVDVDESFTLVHOVRETTLAAHANODVPFE 2937
Qy 12 ----- 11
Db 2938 QIVSNILPGSSDTSRNPVLQMPALHSQQNLGKVRLEGIEEIIISIAETTRFDEIFHLVQ 2997
Qy 12 ----- 11
Db 2998 EAERLNGSIVYAADLFVPETIQSVITIFQGILOKGLGEPMPVASMALDGGLESLSRSTGL 3057

Qy 12 -----VAVGPDV----- 18
||| :|
Db 3058 LHPQQTDPDASVVOIFKQVAVNPDVIAVREDETRLSVADLRKSDQVACHLSRRGIA 3117
Qy 19 ----- 18
Db 3118 PETFVAILAPRSCETIVAILGVLKANLAYLPLDVNVPPASRLEAILSEVSGSMLVGAET 3177
Qy 19 ----- 18
Db 3178 PIPEGMAEATIRITEILADAKTDDINGLAASOPTAASLAYVIFTSGSTRPKGVMEHR 3237
Qy 19 ----- 18
Db 3238 GIVRLTKQTNITSKLPESFHHMHI SNLAPDASVMEVFTLLNGGLVCIDYFTLLESTAL 3297
Qy 19 ----- 18
Db 3298 EKVFQDQVNVALLPPALLKQCLDNSPALVKTLSVLYIGDRLDASDAKAKARGLVQTAQ 3357
Qy 19 ----- 18
Db 3358 NAYGPTENTVMSTIYPIAEDPFINGVPIGHAVNSGAFVMDNQOITPPGAMGELIVTGD 3417
Qy 19 ----- 18
Db 3418 GLARGYTTSLNTGRFINVDIDGEOVRAYRTGDRVRYRPKDLOIEFFGRIDHVKIRGHR 3477
Qy 19 ----- 18
Db 3478 IEPAEVEYALLSHDLVTDAAVVVTHSOENODLEMVGFVAARVADVREDESSNOVQEWTHF 3537
Qy 19 ----- 18
Db 3538 DSIAYADITTIDQOGLGRDFMSWTSMYDGLIKKSOMQEWLDDTMRSLLDSQPPGHVLEV 3597
Qy 19 ----- 18
Db 3598 GTGTGMVLFNLGREGGLOSYYGLEPSPSATAFVYNKAAKSPGLEDRIRVEVGTTATDIDL 3657
Qy 19 ----- 18
Db 3658 GDDLHAGLVVNVSAQYFPSQDYLAQLVRLDTKVPGVVERIFFGDMRSHAINRDFLVARAV 3717
Qy 19 ----- 18
Db 3718 HALGDKATKAEIOREVVRMBESEDLLVDPAPFTSLTTQVENIKHVEILPKRMTRATNELS 3777
Qy 19 ----- 18
Db 3778 SYRYAAVLHVNDLAKPAHKVSPGANVDPFAATKMDRDALIRLLRGTKISDHIAIANIPNSK 3837
Qy 19 ----- 18
Db 3838 TIVERTICESVYDLGGDAKSDNRVSWLSAARSNAVKVASLSAIDLVDIAQEAQFRVEIS 3897
Qy 19 ----- 18
Db 3898 CAROMSONGALDAVFFHHLGSPSSSHVLIDFLTDHQRPEALTNHPLHRAQSRVERQI 3957
Qy 19 ----- 18
Db 3958 RERLQTLIPAYMPIAQIMVLDKLPLNANGKVDKQLTORAQTPKAKOVSAVPAPRTEIE 4017
Qy 19 ----- 18
Db 4018 RVLCOEFSVDLGVDIGIMENFFDLGHSLSMATKLAARISRRLETHSVSKEIFDHPRVCDL 4077
Qy 19 ----- 18
Db 4078 VLIVOQGSAPHDPVISTKYTGVPDQSPFAQGRLLWFLDQNFATWYLMPLAVLRGAMNVH 4137

QY 19 ----- 18
Db 4138 ALTAALLALLERRHELLRTTFYEQNGVGMKNPVVTETLRIIDLSNGDGYLPTLKKQET 4197
QY 19 ----- 18
Db 4198 APFHLETEGMRVALLRGLPGDYLILSVNMHHIISDQWSVDVLFQELGQFYSTAVKGDPL 4257
QY 19 ----- 18
Db 4258 SOTTPLPIHYRDFALWOKPTQSEHEROLQYWEQVDSAPAELLTDLPRPSILSGQAG 4317
QY 19 ----- 18
Db 4318 EMSVTIEGALYKULEEFCRVRHVTFSVLLAALRAAHYRLTGSSEDATIGTPIANRRPEL 4377
QY 19 ----- 18
Db 4378 EQIGFFVNTQIRITVNEDETFESLVQVRSTATAAFAHQDVPEKIVSTLLPGSRDAS 4437
QY 19 ----- 18
Db 4438 RNPLVQLMPAVHSQKNLGEKLENHSEVVPTEITTRFDLEFHLFQDDKLEGSILYSTD 4497
QY 19 ----- 18
Db 4498 LFEAVSQSLLSVFQELIRRLGNGDPVPISTLPLODGIYDLQROGLLDVQKTEYPRDSSV 4557
QY 19 ----- 18
Db 4558 VDFHEQVSINPDSIALIHGSEKLSYAQLDRESDRVARWLRHRSFSSDTLIAVLAPRSC 4617
QY 19 ----- 18
Db 4618 TIIFLILKANLAYLPLDVKAPARIDAIVSSLPKNKILILGANVTPPKLOEAAIDFVP 4677
QY 19 ----- 18
Db 4678 IRDFTTLTDGLODPTIERPSAQSLAYAMFTSGTRPKGMVQHRNIVRLVKNVNV 4737
QY 19 -----FOAHO----- 23
Db 4738 AKQPAARIAHISLAFDASSWEIYAPLLNGGAIVCADIYTTIDPQALQETFOEHEIRGA 4797
QY 24 ----- 23
Db 4798 MLPPLLKQCLVOAPDMI SRLDILFAAGDRFSSVDALQORLVGSGVFNAYGPTENTILS 4857
QY 24 ----- 23
Db 4858 TIYNVAENDSFVNGVPIGSAVNSGAYIMDKNOQLVPAGVMGELVVTGDGLARGYMDPKL 4917
QY 24 ----- 23
Db 4918 DADRFIQLTVNGSEQVRAVYRTGDRVYRPKDFQIEFFGRMDQOIKIRHRIEPAEVEQAF 4977
QY 24 ----- 23
Db 4978 LNDGFVEDVAIVIRTPENQEPWVAFVTAKGDNSAREEATTOIEGWEAHFEGGAYANIE 5037
QY 24 -----EDT----- 26
Db 5038 EIESEALGYDFMGWTSYDGTEDIKOEMREWLNDTMRSLLDGKPAGRVLEVGTTGTGMIF 5097
QY 27 -----ERYV----- 30
Db 5098 NLGRSGLERYIGLEPAPSAEAFVNNAAKSFPGLAGRAEVHVGTAAADVGTLOGLTSDMAV 5157
QY 31 ----- 30
Db 5158 INSVAQYFPTPEYLAETIKSLVQVPMKRIYVLGDMRSMWNRDFAAARAAYSADNASKD 5217
QY 31 ----- 30

Db 5218 RVROKMMELBEKEBELLDVPAFFTALASQLODRIOHQHVEILPKRMKATNELSSRYAAVLH 5277
QY 31 ----- 30
Db 5278 ISDEPLPIYKIDPEAWINFEGSRLTREALAOVLKENENAEVAISNIPYSKTVVERHIVR 5337
QY 31 ----- 30
Db 5338 SLDOEDANAPESMDGSDWISAVRTRAQOCHTSLASDLFDAEDAGFRVSVWARQHSOH 5397
QY 31 -----LTN----- 33
Db 5398 GALDAVFHHLXPATEDSRVLKIFPTDHQGRPLKSLTNOPLLPAPSRRAELLIREGLQTL 5457
QY 34 ----- 33
Db 5458 PPYIMPSQITLIDRMLNANGKVDRELARAKITQSKPVEDIVPPRNSVEATVCKGFT 5517
QY 34 ----- 33
Db 5518 DVLGVEVGITDNFFNLGHSMLATKLAARLGRQLNTRISVRDVPDQVYVADLAIVORNS 5577
QY 34 -----LNTGA----- 38
Db 5578 APHEPIKPADYTGVPVQSFQAQRLWFLDQLVNGATWYMLPLGIRLHLSRVDALATAISA 5637
QY 39 ----- 38
Db 5638 LEORHEPLRTTFHEEDGVGVQVQDHRPKDLRIIDLSTQPKDAYLAVLKHEQTTLFDLAT 5697
QY 39 ----- 38
Db 5698 EPGMRVALIRLGEHEHILSI VMHHIISDQWSVEVLDFEMHRFPYSALRQODPMEQILPLP 5757
QY 39 ----- 38
Db 5758 IQYRFAAQKTEQVQAEHQRLQDLYWTEHLADSTPAELLTDLPRPSILSGRANELPLTIE 5817
QY 39 ----- 38
Db 5818 GRLHDKLRAFRCRVHQATPFVILLAAALRAAHYRLTGAEDATLGTPIANRRNPELENNIGFF 5877
QY 39 ----- 38
Db 5878 VNTQCMRIAIBENDNFESLVRVRSTATSAFANQDVFFESIVSSLLPGSRDASRNPLVQV 5937
QY 39 ----- 38
Db 5938 ILAVHSQODLGLKLTLEGLRDEAVDSAISTRPDVEPHLFEHADRLSGSVLYAKELFKLRTI 5997
QY 39 ----- 38
Db 5998 ESVVSVFLETLRRALDQPLTPLAVLPLTDGVCGEIASKGLDVPRTDYPRDANIVEVFOOH 6057
QY 39 ----- 38
Db 6058 VRATPAIAVKDATSILTIVAQLDQSDRLAIWLSRRHMPETLVGLVAPRSCETIIAMFG 6117
QY 39 ----- 38
Db 6118 IMKANLAYLPDINSAPAARLRSILSAVDGNKVLVILGSGVTAPEQENPEVEAVGICQILAG 6177
QY 39 ----- 38
Db 6178 TGLDKTQGSNARPSATSLAYVIFTSGSTGKPKGMVMEHRSVTRLAKPSNVISKLPOGARV 6237
QY 39 ----- 38
Db 6238 AHLANIAFDASIEIATILLNGATLVCLDYHTVLDCTRLKEVFRESITVTVLMPALLKQ 6297
QY 39 ----- 38

Db 6298 CVABIPETLAHLDLLTYTGGDRVGHGDAMRARSVKIGMFGSGYGTENTVISTIEYVDADE 6357
Qy 39 ----- 38
Db 6358 MFVNGVPIGKTVNSGAYVMDRNOQLVPVGVLVVTGDLARGYTDPSLNKMRFIYT 6417
Qy 39 ----- 38
Db 6418 VNGESIRAYRTGDRVYRPHDLQIEFFGRMDQVKIRGRIEPEGVEVESALLSHNSVODAA 6477
Qy 39 ----- 38
Db 6478 VVICAPADQDSGAEMVAFVAARNTEDTQEEBAVDQVQWETHFETAAYSEVKDIRQSE 6537
Qy 39 ----- 38
Db 6538 VGNDFMGWTSNYDCSEIDKTDHWEHLNDRMILDAPEGHVLGIGTGVMFMENLAKCP 6597
Qy 39 ----- 38
Db 6598 GLQYVGFPEPSKAAQFYNDAAQSPALKDGRSIVHVGATDINKAGPIQPRLVVINSVA 6657
Qy 39 ----- 38
Db 6658 QYFPTPEYLFVFEALVOIPSEVERIVFGDMRTNAINRDFVASRALHTLGEKANKELVRQM 6717
Qy 39 ----- 45
Db 6718 IYELEANEELLLTOPAFTSLRTRLGEKIKHVEILPKTKMATNELSKYRYAAVLHVRGSR 6777
Qy 46 ----- 45
Db 6778 EQSTHIVSPNAWIDFAADGLDROTLINLKEHKDAGTVAIGNTPYSKTIVERFVNKSL 6837
Qy 46 ----- 45
Db 6838 EDDMEEGONSLDGSAAVAVRMAAQCPSLDAMDVKEITAQAGYQVEYSWARQWSONGAL 6897
Qy 46 ----- 49
Db 6898 DAIPHFEPPKEGARTLIEFPTDYEGRVNVTLTNRPLNSIOSRRIGTQIREKLOTLPPY 6957
Qy 50 ----- 49
Db 6958 MIPSRIMVLDQMPVNNNGKIDRKELVRRRAIVAPKPSAATRVAPRNEIEAILRDEFEDVL 7017
Qy 50 ----- 49
Db 7018 GTEVSVDLNFDFDLGGHSLMATKLAARVSRRLDAHISIKDVPDQPVLDLAASIORESAPH 7077
Qy 50 ----- 49
Db 7078 EPIQRPYTPAEQSFQAGRLWFLDQNLGATWYLMPLAIRIGQLRVAALSAAALFALER 7137
Qy 50 ----- 49
Db 7138 RHETLRTTFEESGQVQIVGEARNSDLRVHDVSTGDDGEYLEVLRREQTVPFDLSSEPG 7197
Qy 50 FVHLVK----- 56
Db 7198 WRVCLVKTGEEDHVLIVMHHIYDGMVSDILRGELGQFYSAALRGQDPLLHANPLPIQY 7257
Qy 57 ----- 56
Db 7258 RDPAAQREAKQVEEHQRQLGWSKQLVDSTPAELLTDLPSPILSGRAGSVDTVIEGSV 7317
Qy 57 ----- 61
Db 7318 YGALQSFCTRSTVTFVLLTVFRIAHFRLTAVDDATIGTPIANRNRPELETLVGCFVNT 7377
Qy 62 ----- 61
Db 7378 QCMRISTADDNFEGLVRQVRNVATAAYANQDVPFERIVSALVPGSRNTSRNPLVQLMFA 7437

Qy 62 ----- 61
Db 7438 VOSVEDYDQVRLEGLSVNMPGEASTRFDMEPHLVPGDKLTGSLVYSSDLFEQGTIONF 7497
Qy 62 ----- 61
Db 7498 VDIFQECRLSVLDQPLTPISVLPSNAISNLSLDLLEMTSDYPRDRTVVDLFREQAAI 7557
Qy 62 ----- 61
Db 7558 CPDSIAVKOSSQLTYAQLDEQSDRVAWLHERHMPAESLVGLVSPRSCETIIAYFGIMK 7617
Qy 62 ----- 61
Db 7618 ANLAYPLDYAPDARLAAILDITVEGERLLLLGAGVPGQIQIPRLSTAYIAEALSHATT 7677
Qy 62 ----- 65
Db 7678 VDVTSIPQPSATSLAYVIFTSGTGPKGMIEHGRGIVRLVRDTNVMNVPPESGALPVSH 7737
Qy 66 ----- 65
Db 7738 FSNLAWDAATWEIYTAVLNGGTVVCIDRDTMLDIAALNSTFRXENVRAAFTTAPFLKQCL 7797
Qy 66 --APNITANL----- 73
Db 7798 AETPELVANLEILHTAGDRLDPCDANLAGKTAKGGIFNVLGHNTENTAYSTFYVPVGEETF 7857
Qy 74 ----- 73
Db 7858 VNGVPVGRGISNGSHAYIIDRHOKLVPAGVMGELILTGDGVARGYTDALNKRPFYIDIN 7917
Qy 74 ----- 73
Db 7918 GKSTMSYRTGDKARYRPRDQGLEFFGRMDQWVKIRGRIEPEGEVELTLLDHKSVLAATVV 7977
Qy 74 ----- 73
Db 7978 VRRPPNGDPEMIAFITIDAEDDVQTHKAIYKHLQGLPAYMIPSHLVILDMPVTDNKV 8037
Qy 74 ----- 73
Db 8038 DRKDLALRAQTQVKRRSTAARVPRDEVAVLCEEYSNLLLEVEGITDGFDFLGGHSLA 8097
Qy 74 ----- 73
Db 8098 TKLAARLSRQLNTRSVKDVDPQILADLADIIRRGSHRHDPITPYTGPVEQSFQAGR 8157
Qy 74 ----- 73
Db 8158 LWFLEQLNLGASWYLMPPFAIRMRGLOTKALAVNALVHRHEALRTTFEDHDGVGVQVI 8217
Qy 74 ----- 73
Db 8218 QPKSSQDLRIIDLSDAVDDTAYLAALKREOTTAPDLTSEPGWRVSLRLGLGDDVILSIVM 8277
Qy 74 ----- 73
Db 8278 HHIISDGHVTVLRLQELGQFYSAAIRGQEPLSQAKSLPIQYRDFAVMQROENIQEQAQK 8337
Qy 74 ----- 73
Db 8338 LKYSQQADSTPCFELTDLPRPSILSGEADAVPMVIDGTVVYQLLTDPCRTHQVTSFVSL 8397
Qy 74 ----- 73
Db 8398 LAAPFTHYRLTGLDATVGTPIANRNRPELEGLGFFPWTCQMRMAISETETTESLVQO 8457
Qy 74 ----- 73
Db 8458 VRLTTTEAFANQDVPFEQIVSTLLPGSRDTSRNPVQVMFALQSQOQLGRIOLEGMTDEA 8517

QY 74 ----- 73
Db 8518 LETPLSTRDLVHLFOEVGKLSGSLYSTDLFEVETIRGIIVDVLFLEILRRGLEQPKQRL 8577
QY 74 ----- 73
Db 8578 MAMPITDITKLRDOGLLTAKPAYPRESSVIDLFRQVAAAPDAIAVWDSSTLTLYADL 8637
QY 74 ----- 73
Db 8638 DQOSNLAHWLCORNMAPETLVAVFAPRSCLTIVAFGLVKANLAYPLDVNAPARIEA 8697
QY 74 ----- 73
Db 8698 ILSAVPGHKLVQAHGPELGLTMADELVOIDEALASSSGDHEQIHASGPTATSLAVV 8757
QY 74 ----- 73
Db 8758 MFTSGTGPKGVMIDHRSIIRLVKNSDVVATLPTVRMANVSNLAFDISVQEIYTALLN 8817
QY 74 ----- 73
Db 8818 GGTLVCLDYLLDLSKILYNVFEAQVNAAMFTPVLLKQCLGNMPAIISRLSVLFNVGDR 8877
QY 74 ----- 73
Db 8878 LDAHDAVASGLIODAVYNAYGPTENGQMSTMYKVDVNEPFVNGVPIGRSITNSGAYVMD 8937
QY 74 ----- 73
Db 8938 GNQOLVSPGVMEIVVTGDLARGYTDALDEDRFVHVTTIDGENIKAVRTGDRVRYRPK 8997
QY 74 ----- 73
Db 8998 DFEIEFGRMDQVKIRGRIEPAEVEHALLGHDLVHDAAVLVRKANPEMIAFITSQ 9057
QY 74 ----- 73
Db 9058 EDETIEQESNKVQVQGEHFDVSRYADIKDLDTSTFGHDFLGWTSMDYDGDIPVNEKME 9117
QY 74 ----- 78
Db 9118 WLDETTASLLDNRPBGHILEIGAGTGMILSNLKGVDGLOKYVGLDPAPSAAIFVNEAVKS 9177
QY 79 ----- 78
Db 9178 LPSLAGKARVLVGTALDIGLDKNEIOPELVVINVAQYFTSEYLLIKVYKAVVEVPSVK 9237
QY 79 ----- 78
Db 9238 RVFGDIRSQALNRDFLAARAVRALGDNASKEQIREKIAELESEEBELLVDPAFFVSLRS 9297
QY 79 ----- 78
Db 9298 QLPNIKHEVLPKLMKATNELSYAAVLHISHNEEBEQLLIDIDPTAWDFAATQKDS 9357
QY 79 ----- SVC 81
Db 9358 QGLRNLLQGRDDVMIAGNIPYSKTIHERHIMNSLDQDHVNSLDGTSWISDARSAAC 9417
QY 82 ----- 81
Db 9418 TSPDAPALTOLAKEGFRVELSWARQSONGALDAVFHRLATDANCERSRVLVHFPTDHO 9477
QY 82 ----- 81
Db 9478 GRQLRTLNRPLQARSRRIESQVFEALQALPAYMIPSRIVLPMQPTNANGKVDKQKL 9537
QY 82 ----- 81
Db 9538 ARRAQVAKRKAVSARVAPRNDTEIVLCEEYADILGTEVGITDNFFDMGGHSLMATKLA 9597
QY 82 ----- 81

Db 9598 RLSRRLDTRVTKEVFDKPVLDADLAASIEQGSTPHLPIASSVYSGPVEQSYAQRLWFLD 9657
QY 82 ----- 81
Db 9658 QFNLNATWYHMSLAMRLLGPLNMDALVALRQBORHETLTTTPEAKDQIGVQVVHEAGM 9717
QY 82 ----- 81
Db 9718 KRLKVLDSLKNEKEHMAVLENEQMRPFTLASPEQWKHGLARLGPTEYILSLVMHHMPSD 9777
QY 82 GWS ----- 84
Db 9778 GMSVDILROELGQFVSAALRGDRDPLSQVKPLPIQYRDPFAAWOKENAAQVAEHERQLAYWEN 9837
QY 85 ----- 84
Db 9838 QLADSTPGELLTDPRPQFLSGKAGVIPVTIEGPVYEKLLKFSKERQVTLFSLVLLTAFRA 9897
QY 85 ----- 84
Db 9898 THFRLTGAEDATIGTPIANRNPELEHIIGFFVNTQCMRLDLDGTSTFESLVQHVRSVAT 9957
QY 85 ----- 84
Db 9958 DAYSQDIPFERIVSALLPGSRDASRSLIQLMPALHQSOPDLGNITLEGHERLPTSVA 10017
QY 85 ----- QTIN ----- 88
Db 10018 TRFDMFHLFQEPNKLSGSILLFADELFOPETINSVVTVFQEILRRGLDQPOVSISTMPLT 10077
QY 89 ----- 88
Db 10078 DGLIDLEKLGLEIESSNPPRDYSVVDVFRQOVAANPNAPAVVDSETSMYSYTLQKSEO 10137
QY 89 ----- PE ----- 90
Db 10138 IAAMLHAQRLRPESLVCNAPRSFETIVSLFGILKAGYAYLPDVNSPAARIQPLSEVE 10197
QY 91 ----- 90
Db 10198 GKRLVLLSGIDPMOSDRMDVETARIQDILTNTKVERSPPMSRPSATSLEYVFTSGSTG 10257
QY 91 ----- 90
Db 10258 RPKGVMIHRNLRVLKQSNVTSQLPQDLRMAHISNLPADASIWEIFTAILNGGALICID 10317
QY 91 ----- 90
Db 10318 YFTLLDSQALRTTFEKARVNATLFPALLKECLNHAPTLFEDLKVLYIGGDBLDATDAAK 10377
QY 91 ----- 90
Db 10378 IQALVKGTYVNAYGPTENTVMSTIYRLTDGESYANGVPIGNVAVSSSGAVIMDOKORLVPP 10437
QY 91 ----- 90
Db 10438 GVMGELVWSDGLARGYTNSTLNADRFVDIVINDOKARAYRTGDRTRYRPKDGSIEFFGR 10497
QY 91 ----- 90
Db 10498 MDQOVKIRHRVEPAEVEQAMLGKAIHDAAVVQAVDQOQETEMIGFVSNASDRPSEGE 10557
QY 91 ----- DOT ----- 93
Db 10558 EITNOVQEMEDHFESTAYAGIEAIDQATLGRDFTSMTSMYNGNLIDKAEMEELDDTMS 10617
QY 94 ----- 93
Db 10618 LLDKEDARPCAEIGTGTGMVLFNLPKNDGLESYVIEPSRSAAALFVDKAAQDPFLOGKT 10677
QY 94 ----- 93

Db 10678 OILVGTAEIKLVKDFHPD VVINSVAQYPPSRSYLVQIASELIHMTSVKTIFFGDMRSW 10737
Qy 94 -----DP----- 95
Db 10738 ATNRDFLVSRALYTLGDKATKQDIQREVARLEENEDELLVDPAPFTSLTSQWPGKVXHE 10797
Qy 96 ----- 95
Db 10798 ILPKMRMSTNELSSRYAAVLHICRDGEGRNRYGRRVHSVEENAWIDFASGMDRHALVQ 10857
Qy 96 ----- 95
Db 10858 MLDERRDAKTVAINI PHSNTINERHFTTSLDTEGEGIAQDSLDSAWQSAWKAAWARP 10917
Qy 96 ----- 95
Db 10918 CLSVTELVEIGQAAGFRVSVSWARQSRQHGLDVVFHLEDDRVRVLINPPTDFERLPP 10977
Qy 96 ----- 95
Db 10978 STGLTSRPLORIONRRFESQIREOLOTLPPYMPVPSRIVVLERMPLNANSKVDKRELARK 11037
Qy 96 -----GH----- 97
Db 11038 ARTLOTIKPSATRVAPRNDIEAVLCDFEQAVLGVTVGMDNFFELGGHSLMATKLAARLS 11097
Qy 98 -----ADLV----- 101
Db 11098 RRLDTRSVKDI FNOPILODLADVVTGSGAPHEAIPSTPYSGPVEQSFQGRFLWFLDQLN 11157
Qy 102 ----- 101
Db 11158 LNASWYHMLASRLRGPLRIEALOSALATIEARHESLTTTFEEOGVPVQIVRAARNKQL 11217
Qy 102 ----- 101
Db 11218 RIIDVSGTEDAYLAALKQEQDAAFDLTAEPGRVALLRLGPDHVLVSIWMHHIISDGWSV 11277
Qy 102 -----LY----- 103
Db 11278 DILROELGOLYSNASSQAPLPIQYRDFAIWQODSQIAEHQKQYNWKQLVNSKPAEL 11337
Qy 104 ----- 103
Db 11338 LADFTPRKALSGADVPIEIDDOYQNLRSFCRARHVTFSVALLAARAAHYRLTGAED 11397
Qy 104 ----- 103
Db 11398 ATIGSPIANRNPPELEGLIGCFVNTQCLRIPVKSEDTFTLVKQARETATEAQDNQDVPF 11457
Qy 104 -----ITRFDLEL----- 111
Db 11458 ERIVSSMVASSRDTSRNPLVQWFAVSHQDLGNIRLEGVEGKPVSMASSTRFDEAMHLF 11517
Qy 112 ----- 111
Db 11518 EDQGLGNGVWFSKDLFESETIRSVAVFQETLRLRGLANPHANLATLPLTDGLPSLSLC 11577
Qy 112 -----PD----- 113
Db 11578 LQVNPQDYPDRASVIDVFEQVASIPKSTAVIDAASSQLTYTELDERSSQATWLRQVTV 11637
Qy 114 -----GNQV----- 118
Db 11638 PEELGVLA PRSCETIIAFLGIKANLAYLPLDVNAPAGRIETIILSSLPFGNRLILLGSDT 11697
Qy 119 -----R----- 119
Db 11698 QAVKLHANSVRFRTRISDALVESGSPTEELSTRPQAQSLAYVMFTSGTGVPKGVMEHR 11757
Qy 120 GVT----- 122
Db 11758 GITRLVNSNVAKOPAAAAIAHLSNIAFDASSWEIYAPLLNGGTWVCIDYVTTIDIKAL 11817

Qy 123 ----- 122
Db 11818 EAVFKQHHRGAMLPALLKQCLVSAPTWMISSEILFAAGDRLLSSODAILARRAVGSGVY 11877
Qy 123 ----- 122
Db 11878 NAYGPTENTVLSTIHNIGENAEFNGVPIGNAVSNGSAFMDQONQOLVAGVIGELVVTG 11937
Qy 123 ----- 122
Db 11938 DGLARGYTDKLRVDRFIYITLDGNRVRAYRTGDRVHRPKDQGIIEFFGRMDOQIIRGH 11997
Qy 123 ----- 122
Db 11998 RIEPAEVEQALARDPAISDSAVITQLTDEEPELVAFPSLKGNGANGTNGVNGSDQEKID 12057
Qy 123 ----- 122
Db 12058 GDEQHALLMENKIRHNQLALLPTYMIPSRIIHVDOLPVNANGKIDRNELAVRAQATPRTS 12117
Qy 123 ----- 122
Db 12118 SVSTYVAPRNDIETIICKEFADILSVRVGITDNFFDLGGHSLIATKLAARLSRRLDTRVS 12177
Qy 123 ----- 122
Db 12178 VRDVFDPVVGQLAASIQQGSTPHEAI PALSHSGPVQQSPAQGRFLWFLDRFNLAAMYIM 12237
Qy 123 ----- 122
Db 12238 PFGVRLRGPLRV DALQ TALRALEERHELLRTTPEEQDGVGMQIVHSPMRDICI VWDISGA 12297
Qy 123 ----- 122
Db 12298 NEDLAKLEBQAPFNLSTEVAMRV ALFKAGENHHILSI VMHHIISDGWSVDI PQBELAQ 12357
Qy 123 ----- 122
Db 12358 FYSVAVRGHDP LSOVKPLPIHYRDFAVWQRQDKQAVHESQLOQWIEQLADSTPAELISD 12417
Qy 123 ----- 122
Db 12418 FNRPEVLGSEAGTVPIVIEDEVYEKLSLFCRNHQVTSFVLLAARVAVHYRLTGAEDATI 12477
Qy 123 ----- 122
Db 12478 GTPIANRNRPELEDLIGFFVNTOCMRIALEEHDFLSVVRVRSTAASAFENQDVPFERL 12537
Qy 123 ----- 122
Db 12538 VSALLPGSRDASRNPLVQLMFVVHVSQRNLKQLGLEGEGBTPYTTATTRFDEPHLFEQD 12597
Qy 123 ----- 122
Db 12598 KGLAGNVVFAADLFEAATIRSVVEFHEILRRGLDQDPDIAISTMPLVDGLAALNSRLPA 12657
Qy 123 ----- 122
Db 12658 VEDIEPDFA TEASVVVVFQTVQVNPALAVTDTSTKLTVAELDQOSDHVAWLSKOKLP 12717
Qy 123 ----- 122
Db 12718 AESIVVVLA PRSSETIVACIGILKANLAYLPMDSNVPEARQAILSEIPGEKPVLLGAGV 12777
Qy 123 ----- 122
Db 12778 PIPDNKTADV RMVFI SDIVASKTDSYSGTPRSASSLAYVIFTSGTGRPKGVMEHRG 12837
Qy 123 ----- 122
Db 12838 VISLVKQNASRIPOSLRMAHVSNI LAFDASVWEIIFTLLNGGTLFCISYFTVLDSKALSAA 12897

QY 123 ----- 122
Db 12898 FSDHRINITLLPALLKQCLADAPSVLSLESYIGGDRDLGDADATKVKDLVKGKAYNAY 12957
QY 123 ----- 122
Db 12958 GPTENSVMSTIYTIETHETFANGPIGTSLGPKSKAYIMDQDQLVPAGVMGELVWAGDGL 13017
QY 123 ----- 122
Db 13018 ARGYTOPSLNTGRFINHTIDGKQVARYTGDRVRYRPRDYQIEFFGRLDQOIKIRHRIE 13077
QY 123 ----- 122
Db 13078 PAVEQALLSDSSINDAVVVSQAQKGLEVMGYITQAAQSVDKKEASNKVQWEAHFDS 13137
QY 123 ----- 122
Db 13138 TAYANIGGIDRDALGDQFLSWTSMYDGLIPREEMQEWLNDTMRSLDNPQPGKVLIGT 13197
QY 123 ----- 122
Db 13198 GTGMVFLNLGKEGLOSAGLEPSRSVTAMWNKAIEFPFSLAGSARVHVGTAEIDISSDG 13257
QY 123 ----- 122
Db 13258 LRSDLVINSVAQYFFPSREYLAEALTANLIRLPGVKRIFFGDMRTYATNKDFLVARAVHTL 13317
QY 123 ----- 122
Db 13318 GSNASKAMVQVAKLEDDDEELLVDPAFTSLSDQFPDEIKHVEILPKRMAATNELSSY 13377
QY 123 ----- 122
Db 13378 RYAAVHVGHGHPNGEDEDKQAVKINPKAWDFAGTRMDRQALLQLLQDRGDDVV 13437
QY 123 ----- 122
Db 13438 AVSNIPYSKTIEMRHLSQSLLDDEDTGTSVDGTAWISRTOSRAKECPALSVADLIEIGK 13497
QY 123 ----- 122
Db 13498 IGFEVEASWARQHSQRGLDAVFRHPPRHSRGHVMPFRFTEHKGRSSSLTNRLPLHLQ 13557
QY 123 ----- 122
Db 13558 SRLEAKVRERLOSLLPPYMIPSRITLLDQMLPTNSGKVDKRLARQARVIPSAASTLD 13617
QY 123 ----- 126
Db 13618 FVAPRTEIEVVLCEEFTDLGLGVKVGITDNFFELGGHSLLATKLSARLSRRLDAGITVKQV 13677
QY 127 ----- 126
Db 13678 FDQVPLADLAASILQSSRRHSRPSLPYEGPVEQSAQRLWFLDQFNIDALWYLIPFAL 13737
QY 127 ----- 126
Db 13738 RMRGFLQVDALAAALVALEERHESLRTTFRDGVGIQVQVPLRTTKDIRIIVSGMRDD 13797
QY 127 ----- 132
Db 13798 DAYLEPQKEQOTPFDLASEPGWRVALLKLGKDDHILSVVHHIISDGNSTEVLORELQ 13857
QY 133 ----- 132
Db 13858 FYLAASKGAPLSQVAPLPIQYRDFAVMQROEQVAESQRLDVKWKQADSPPELLAD 13917
QY 133 ----- 132
Db 13918 YTRPNVLSGEAGSVFVINDSVYKSLVFCRSRQVTTFTTLAAFPRAAHYRMTGSDDATI 13977
QY 133 ----- 136

Db 13978 GTPIANRNRPELENLIGCFVNTQCMRITIGDDTFESLVQVVRSTTATAFENQDVPPFERI 14037
QY 137 ----- 144
Db 14038 VSTLSAGSRDTSRNPFLVQLLFAVHSQOGLGRIQLDGVVDPEVLSTVSTRFDLEFHAFOEA 14097
QY 145 ----- 144
Db 14098 DRLGNSVMFATDLFOPETIQGFVAVVEVLQRLGEPQSPQSIATMPLAEGIAQLRDAGALQ 14157
QY 145 ----- 144
Db 14158 MPKSDYPRNASLVDVFQQAAMASPSTVAVTDSKLTLYAELDRLSQQAASYLRRQOLPAE 14217
QY 145 ----- 144
Db 14218 TMVAVLAPRSCETIIAFLAILKANLAYMPLDVNTPSARMEALISSVPGRRLLVLVSGVRH 14277
QY 145 ----- 144
Db 14278 ADINVPNAKTMLISDVTGTDAIGTPEPLVVRPSATSLAYVFTSGSTGKPKGMVMEHRA 14337
QY 145 ----- 144
Db 14338 IMRLVKDSNVVTHMPATRMHAHVNTNIAFDVSLFEMCATLLNGTGLVCIDYLLDSTMLR 14397
QY 145 ----- 144
Db 14398 ETFEREQVRAAIFPPALLRQCLVNMPPDAIGMLEAVYVAGDRFHSRDARATQALAGPRVYN 14457
QY 145 ----- 144
Db 14458 AYGPTENAILSTIYNIDKHDPYVNGVPIGSAVNSGAYVMDRNQQLPPGVGELVVGTGE 14517
QY 145 ----- 144
Db 14518 GVARGYTDASLTDREVTVTIDGQORAYRTGDRVYRPRKGFQIEFFGRLDOOAKIRGHR 14577
QY 145 ----- 144
Db 14578 VELGEVEHALLSENSVTDAAVVLRTWEEDPOLVAFVTTDHEYRSGSSNEEDPYATQAA 14637
QY 145 ----- 144
Db 14638 GDMRKELRSLLPYVMVPSRVITILRQMPLANGVKDRKDLARRAQMTPTASSSGPVHVAPR 14697
QY 145 ----- GVT 147
Db 14698 NETEAAICDEFETILGVKVGIT 14719

RESULT 3

T30192

probable peptide synthetase - Aureobasidium pullulans

C:Species: Aureobasidium pullulans

C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000

C:Accession: T30192

R:Peery, R.B.; Thornevell, S.J.; Tobin, M.B.; Skatrud, P.L.

submitted to the EMBL Data Library, January 1997

A:Description: Discovery of an MDR-like gene adjacent to a peptide synthetase in Aure

A:Reference number: Z20767

A:Accession: T30192

A>Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-10797 <PEE>

A:Cross-references: EMBL:U85909; NID:G4099310; PID:G4099313; PIDN:AAD00581.1

C:Genetics:

A:Introns: 2078/3; 4142/3; 6075/3; 7963/3; 7985/2

C:Superfamily: acyl carrier protein homology

C:Keywords: carrier protein

F;1618-1688/Domain: acyl carrier protein homology <ACP1>

F;3682-3752/Domain: acyl carrier protein homology <ACP2>

F:5615-5685/Domain: acyl carrier protein homology <ACP3>
F:7503-7573/Domain: acyl carrier protein homology <ACP4>
F:9683-9752/Domain: acyl carrier protein homology <ACP5>

Query Match		31.6%; Score 244; DB 2; Length 10797;
Best Local Similarity		1.1%; Pred. No. 55;
Matches		94; Conservative 27; Mismatches 27; Indels 8749; Gaps 30;
QY	1 AAG-----	3
DB	575 AAGLSLRAIQVASQLRQGCADLCALLESVNLDELVRKADENINDNDSEKWK	634
QY	4 -----	3
DB	635 QTVHELRSVERDPESKAFVSGIEDVLPCTPLQDAMLVETAKRPOAYCNELRLLTVSPKIP	694
QY	4 -----GILH-----	7
DB	695 VERVQALFALAQRHTALRSQFMPSGVSHCAVTVQVIMKTLVTSQFAHVKSFTTCGWSVNR	754
QY	8 -----LELIVA-----	13
DB	755 ETLRLPLHFQYKCGAEAEILVAIHVHALYDQMSVEVILELTLQNERTPERPSFGAVN	814
QY	14 -----	13
DB	815 KFFNLRRSEDQSHLDFWGEYLSVDTVPGRLNLSPPMMPPPOPLQSIQHTIEMDMETLROA	874
QY	14 -----VGP-----	16
DB	875 AHSYSCSAHVFOAYAILGFYMGTEDTVEGTFSGRTLPIVEIESVGVPLLSLTPRI	934
QY	17 -----DVQAHQED-----	25
DB	935 NTLESRKFSVLSRLQEDNRKIMRHSMTSLADIKKACGFNGEAVFDSIFVWQETARPD	994
QY	26 -----	25
DB	995 RAQTLNLVLAHDYLEFNLTLLEPTQOGVTKATYQSSLLPLQHVKTLLQOLDALXIV	1054
QY	26 -----	25
DB	1055 VARPETHNNEISDQLPISVLSVANSEPOSFVYKAGLSGVNHALNNSGGLALVFAHIR	1114
QY	26 -----	25
DB	1115 ECTSRMESLTYGELNTRANQLANYLISQAKRDELICVMEKSVSLYLILAAVKAAGCY	1174
QY	26 -----	25
DB	1175 LPLVPETPAARIRQILAEADVVKFCILTDSSMAPVIAVSRCHIMNVDTTDCSAQSGTGPQL	1234
QY	26 -----TERYVLTNLNI-----	36
DB	1235 DFKPTDIAVAVTSGTTGKPKGVLTQENILSNLEVLKSIYVPVPEGSRLQACNQAFDVS	1294
QY	37 -----	36
DB	1295 VFEIFFTYTGMCLCSASKVDVFRDPEKANEILEITHLSLTPVAALTDPAHPRVKFLV	1354
QY	37 -----	36
DB	1355 TAGEAVTHHVHGAWAKGLYQGYGPSETTNICTVNSAVESDHVINNIGPAFENTSFAVLT	1414
QY	37 -----	36
DB	1415 QGDDFQVPLGGLGELFCGQQVFRGYQNMPELTESKIINHFNPGRIYRSGDLGRLLPDG	1474
QY	37 -----	36
DB	1475 TILIOGRTDDORKIRGORIELGEISGCLLQFPSONCAIEVIKTADEKRLMAFWIPSGYS	1534
QY	37 -----	36

DB	1535 KDSYSILQPDKNLEEIIKSIYAHADNLPAYMVPDALVPVSAIPOTSQOGKIDKRRRLASDG	1594
QY	37 -----	36
DB	1595 SALTVEDLNAYSRGADDDTSELSATEQOALASALADTLOWQTSIGRSTISFFALGLDSVS	1654
QY	37 -----	36
DB	1655 AIRLATNLRKEYGYSIDVSQILKRPTIARLAPLLGGESSKOTNEPVTADCEAAVGSYLHD	1714
QY	37 -----	36
DB	1715 ESVVSQHLHEGQTVSQVLPCTPLQEAMLSARDTSGSSAYRNKTLFSLHGSVDKLKACWEV	1774
QY	37 -----	36
DB	1775 MLORHDILRTIFLSTEDSRPPFVOAVLSQWTLPWQECDDIPDQLSTLLDSAKAGDSIVD	1834
QY	37 -----	36
DB	1835 HSPPWKIQVYRSESTVYLLDDMHHALYDANAMSNLLYEVEQLYKDQSLSAVSPFKPFLNF	1894
QY	37 -----	36
DB	1895 MISTSVEEADALPDQLREFVPKPKRTDVKSGFGTITGRNLNYSPKMVETFLSKHSTML	1954
QY	37 -----	36
DB	1955 SLTQAMMMKTLAASQSYSDVCCGVNVSGRSPVVDGIESLVAPCENITPVRVDLSKHSNL	2014
QY	37 -----	36
DB	2015 GLVKALQVNIQSLPYQLTPLRRIQAAGTNGKRLPDSLVLLOQDITDLSAIWRLEGES	2074
QY	37 -----GAELLRD-----	43
DB	2075 GVMDDETSLNPLPLLEGPELHLLISGPETSGTALEYLDADGKTITLSYLDLHRRS	2134
QY	44 -----	43
DB	2135 EHLAWRLIEVSRHISRKDLGIVPIYIPQCISLYISQALIKSGAANFCPLNDVPGERLKP	2194
QY	44 -----	43
DB	2195 ILKDTASILLTTAAMPKLPLEGITVIIVDDDELSEFKSDREPAGEKITSPDTSLAYI	2254
QY	44 -----	43
DB	2255 MYTSGSTGLPKAVCLSHRAVTQSLLAHDFIPSPSRFLOFASPTFDVSFFPPMYRGA	2314
QY	44 -----	43
DB	2315 TLVSVERNRLGDLPGTITSLNIDAAELTPSVAASLVRRHRENVPTLRALLTIGEMLNTQV	2374
QY	44 -----	43
DB	2375 IQFGGSGIKSGILYGMYPTEAAIHCTLOPGFGVDLPAGTIGIPLDVTSCFIVKPTST	2434
QY	44 -----PSLGAQFR-----	51
DB	2435 KHASOLEILPIGEIGELVIGGHQADGYNREOTRAAFVTHPKFGGLYRTGDKARLHRN	2494
QY	52 -----	51
DB	2495 GTLECYGRISSGQVKLRQORVELGEIEHAASKAGGCHAVIASVISGLLVLCIGDPHRVS	2554
QY	52 -----	51
DB	2555 SKDIKSACQKWLPAYMIPSDIVLLDDFPYLPSPGVKKRLETDYNSNTAQHVSGSSDLSE	2614
QY	52 -----	51

Db 2615 NAREITRIIESVLGVSIDHSTLSDSAAGLSRAIQVASQLRRQCADLGALELLSVSNL 2674
Qy 52 -----VH----- 53
Db 2675 ALDELVRAKADESNINDNDSEKWKQTVHELRSVERDPESKAFVSGIEDVLPCTPLQDAM 2734
Qy 54 ----- 53
Db 2735 LVETAKRPOAYCNELRLTVSPKIPVERVRQALFALAQRHTALRSGFMPGSHCAVTVQVI 2794
Qy 54 ----- 53
Db 2795 WKTLVTSQFAHVKSFTTGWSVTNRETLLRPLHFQYKCSGAEAEILVAIHALYDQMSVEV 2854
Qy 54 ----- 53
Db 2855 ILEDETLLONERTPERPSFGAVNKFNLRRSEDOTSHLDFWGEYLSDVTPGRLPNLSPK 2914
Qy 54 ----- 53
Db 2915 MMPPOQLSQIHTIEMDMETLROAAHSYSCSAHVFFQAAVAAILLGFYMGTEDTVFGTVFS 2974
Qy 54 ----- 53
Db 2975 GRTLPIVEIESVMGPLLSTLPTRLTLESRRKFSVLRSLOEDNRKIMRHSMTSLADIKKA 3034
Qy 54 ----- 53
Db 3035 CGFNPGEAVFDSIFVWOETARPDAQOTLLNLVEAHDYLFENLTLELEPTQOGVKTKATY 3094
Qy 54 -----LVQWVI----- 59
Db 3095 QSSLLPLQHVKTLLQQLDALVKIVVARPETHMNEISDQLPISVLSVANSEPSQSVYKAGL 3154
Qy 60 ----- 59
Db 3155 GSVLENHALNSSLALVFAHDIREGTSRMESLYGELNTRANQANLYLSQGAKRDELI 3214
Qy 60 -----LTE----- 62
Db 3215 CVCMEKSVSLYLSILAAVKACGGYLPVLPETPAARIQILAEADVFKCLTDSSNAPVIAD 3274
Qy 63 ----- 62
Db 3275 VSRCHINVDTTDCSAQSGTGPQLDFKPTDIAYAVFTSGTTGPKPKVLVTQENILSNLEV 3334
Qy 63 -----PEGA----- 66
Db 3335 LSKIPVPEGSRLLOACNQAFDVSVEIFFTWYTMCLCSASKDMVFRDFEKAINELEIT 3394
Qy 67 -----PNITANL----- 73
Db 3395 HSLTPTVAALTDPAHIPPVKFLVTAGEAVTHHVHGWAGKLYQGYGPSETTNICTVNS 3454
Qy 74 ----- 73
Db 3455 AVESDHVINNIGPAPENTSFAVLTOGDDFQLVPLGLGELCFGQGVQVPRGYQNNPELTES 3514
Qy 74 ----- 73
Db 3515 KIINHPNYGRIYRSGDLGRLLPDGTILIQGRTDQDKIRGORIELGEISGCLLOFPSPVQN 3574
Qy 74 ----- 73
Db 3575 CAIEVIKTADKERLMAFWIPSGYSKDSYSILOPKNLEBIIKSIYAHLADNLPAVMYVDA 3634
Qy 74 ----- 73
Db 3635 LVPVSAIPOTSQKIDKRRLASDGSALTVEDLNAYRGADDDTSELSATEQOLASALAD 3694
Qy 74 ----- 73
Db 3695 TLQMSQTSIGRSTSPFALGLDSVSAIRLATNLNRKEYGYSIDVSQILKRPTTARLAPLLGG 3754

Qy 74 ----- 73
Db 3755 ESSKQTNBVTADCEAAVGSYLHDESVVSQLHEHGTQTVSQLPCTPLOEAMLSARDTSGS 3814
Qy 74 ----- 73
Db 3815 SAYRNKTLPSLHSGVDKXKACWEVLMQRHDIILRTIFLSTEDSRFPFQVAVLSQWTLPHQOE 3874
Qy 74 ----- 73
Db 3875 CDDIPDLSTLLDSAKAGGDSIVDHSPPMKIQVYRSESTVYLLLDHMHALYDANAMSNLL 3934
Qy 74 ----- 73
Db 3935 YEVEQLYKDOQSLSPVSPKPFPLNPMISTSVBEADALFRDQLREFVPKPKRTDVKSGFGT 3994
Qy 74 ----- 73
Db 3995 ITGRLNYSKQWVETFLSKHSTTMLSITQAMWMTLASOSYSVDCGCVVSGRSVPVDGI 4054
Qy 74 ----- 73
Db 4055 ESLVAPCFNTIPVRVDLSKHSRNLGLVKALQRWNIDSLPVQLTPLRRIQAQAGTNGKRLF 4114
Qy 74 ----- 73
Db 4115 DSLVLLQDDTTDLDSAIMRLEGESGYMDMRPKLPELEGITIVVDDDELSEFKSDREPAGE 4174
Qy 74 -----TSS----- 76
Db 4175 KITSPDTSLAYIMYTSGTGLPKAVCLSHRAVTSQLLAHDRIIPFSRFLQFASPTFDV 4234
Qy 77 ----- 76
Db 4235 SVPEIFFPWRGATLVSVERNRLGDLPGTITSLNIDAAELTPSVAASLVHRHENVPTLR 4294
Qy 77 ----- 76
Db 4295 ALLTIGEMLNTVIOQFGSICKSGILYQMYGPTAAIHCTLOPGFVDLPAGTIGIPLD 4354
Qy 77 ----- 76
Db 4355 TVSCFIVKTESTKHASOLEILPIGEIGELVIGGHQADGLYNRBEOTRAAFVTHPKFGG 4414
Qy 77 ----- 76
Db 4415 LYRTGDKARLHRNGTLECYGRISSGQVKLRQORVELGEIEHAASKAGGCHAVIASVISGL 4474
Qy 77 ----- 76
Db 4475 LVLFICIGDPRVSSKDIKSACQKMLPAYMIPSDIVLLDDFPYLPSPGVKVKRLETDYNSN 4534
Qy 77 ----- 76
Db 4535 TAOHVSGSSDLSENAREITRIETSVLGVSIDHSTDLSAAGLDSLRAIQVASQLRROSCAD 4594
Qy 77 -----LLSV----- 80
Db 4595 LGALELLSVSNVLDELVRAKADESNINDNDSEKWKQTVHELRSVERDPESKAFVSGI 4654
Qy 81 ----- 80
Db 4655 EDVLPCTPLODAMLVETAKRPOAYCNELRLTVSPKIPVERVRQALFALAQRHTALRSGFM 4714
Qy 81 ----- 80
Db 4715 PSGVSHCAVTVQVIWKTLVTSQFAHVKSFTTGWSVTNRETLLRPLHFQYKCSGAEAEILVA 4774
Qy 81 ----- 80
Db 4775 IHHALYDQMSVEVILEDETLLQNETPERPSFGAVNKFNLRRSEDOTSHLDFWGEYLS 4834

Qy 81 ----- 80
Db 4835 DVTGRLPNLSPKMPPOPLQSIQHTIEMDMETLROAAHSYSCSAHVFFQAAVAILLGFY 4894
Qy 81 ----- 80
Db 4895 MGTEDTVFGTVFSGRTLPIVEIESMVGPLSLTLPRINTLESRKFSVDLSRLQEDNRKIM 4954
Qy 81 -----CG-----WSOTINPE----- 90
Db 4955 RHSMTSLADIKKACGPNPCEAVFDSIFVQETARPDAQAQTLLNLVEAHDYLEFNLTLEL 5014
Qy 91 ----- 90
Db 5015 EPTQGVKTKATYQSSLLPLQHWKTLQOLDALVKIVVARPETHMNEISDQLPISLSVA 5074
Qy 91 ----- 90
Db 5075 NSEPOSFYKAGLSLVENHALNNSGGIALVFAHDIREGTSRMESLTYGELNTRANQLAN 5134
Qy 91 -----DO----- 92
Db 5135 YLISQAKRDELICVMEKSVSLYLILAAVAKGCGYLPVPETPAARIRQILAEADVCF 5194
Qy 93 ----- 92
Db 5195 CLTDSSMAPVIADVSCHIMNVDTTDCSAQSGTGPQLDFKPTDIAYAVFTSGTTGPKGV 5254
Qy 93 ----- 92
Db 5255 LVTQENILSNLEVLKIVPVEGSRLLQACNOAFVDSVFEIFFTWYTCMLCSAKDVNF 5314
Qy 93 -----TDPGH----- 97
Db 5315 RDFEKAINELEITHLSLPTVAALTDPAHI PRVKELVTAGEAVTHVHGAWAGKLYQGY 5374
Qy 98 ----- 97
Db 5375 GPSETTICTVNSAVESDHVINNIGPAFENTSFAVLTOGDDFQLVPLGGLGELCFGGQOV 5434
Qy 98 ----- 97
Db 5435 FRGYQNPPELTESKINHPNYGRIYRSBGDLGRLLPDGTLIOGRTDDQDKIRGQRIELGE 5494
Qy 98 ----- 97
Db 5495 ISGCLLOPFSVQCAIEVIKTADKERLMAFWIPSGYSKDSYSILOPKULEBIIKSIYAH 5554
Qy 98 ----- 97
Db 5555 LADNLPAYMVPDALVPVSAIPOTSQKIDKRRLASDGSALTVEDLNAYSRGADDETSEL 5614
Qy 98 ----- 97
Db 5615 SATEQOLASALADTLQMSQTSIGRSTSPFALGLDSVSAIRLATNLKKEYGYSIDVSQILK 5674
Qy 98 ----- 97
Db 5675 RPTIARLPLLGESSKOTNEPVTADCEAAVGSYLHDESWSQLHGHGOTVSVQLPCTPL 5734
Qy 98 ----- 97
Db 5735 QEAMLSARDTSGSAYRNKTLFSLHGSVDKLKACNEVMLQRHDLRTIFLSTEDSRFPVP 5794
Qy 98 ----- 97
Db 5795 QAVLSQWTLPWQECDDIPDQLSTLLDSAKAGGDSIVDHSPPMKIQYRSESTVYLLLDHMH 5854
Qy 98 ----- 97
Db 5855 HALYDANAMSNLLYEQLYKQSLAPVSKPFLNFMISTSVEEADALFRDQLREFVPK 5914
Qy 98 ----- 97

Db 5915 PFKRTDVKSGGTITGRNLNYSKMWETFLSKHSTTMLSLTOAMMMKTLAASOSYSDVCCG 5974
Qy 98 ----- 97
Db 5975 NVVSGRSPVDGIESLVAPCFNTIPVRVLSKHSNGLVKAQORVNIIDSLPYQLTPLRR 6034
Qy 98 ----- 97
Db 6035 IOAAGTNGKRLPDSLVLLODDTTDLDSAIWRLEGESGVMDMYTSGTGLPKAVCLSHRA 6094
Qy 98 ----- 97
Db 6095 VTOSLLAHRFIPIFSRFLQFASPTFDVSVEIFFPWRGATLVSVERNLLGLDLPCTIT 6154
Qy 98 ----- 97
Db 6155 SLNIDAAELTPSVAASLVHRHENVPTLRALLTIGEMLNTQVIOQFGSGICKSGLYGMYG 6214
Qy 98 ----- 97
Db 6215 PTEAAIHCTLOPGFVDLPAGTIGIPLDTVSCFIVKPTSTKHASQLEILPIGEIGELVI 6274
Qy 98 ----- 97
Db 6275 GGHQADGYLNREEQTRAAVTHPKFGLYRTGDKARLHNGTLECYGRISSGOVKLRGQ 6334
Qy 98 -----A 98
Db 6335 RVELGEIEHAASKAGGCHAVIASVISGLLVLCIGDPHRVSSKDIKSAKOKMLPAYMIPS 6394
Qy 99 DLVLY-----ITRF----- 107
Db 6395 DIVLLDDFPYLPBGKVDKRLKLETDYNSNTAQHVSGSSDLSENAREITRIIESVLGVSIDH 6454
Qy 108 ----- 107
Db 6455 STDLSAAGLDSLRAIQVASQLRRQGCADLGALELLSVSNVLDELVRKADSNINDND 6514
Qy 108 ----- 107
Db 6515 SEKNQOTVHELRSVERDFESKAFVSGIEDVLPCTPLQDAMLVETAKRPOAYCNELRLTV 6574
Qy 108 ----- 107
Db 6575 SPKIPVERVROALPALAQRHTALRSFGMPSGVSHCAVTVIWKTLVTSQFAHVKSFTTGW 6634
Qy 108 -----DLE----- 110
Db 6635 SVTNRETLLRPLHFQYKCSGAEBILVAIHHALYDQWSVEVILEDLTLQWERTPERPS 6694
Qy 111 ----- 110
Db 6695 FGAVNKFNLRRSBDQSHLDFWGEYLSLVTPGRLPNLSKPMPPQPLQSIQHTIEMDM 6754
Qy 111 ----- 110
Db 6755 TLROAAHSYSCSAHVFFQAAVAILLGFYMGTEDTVFGTVFSGRTLPIVEIESMVGPLLSL 6814
Qy 111 ----- 110
Db 6815 LPTRINTLESRKFSVDLSRLQEDNRKIMRHSMTSLADIKKACGPNPCEAVFDSIFVMOET 6874
Qy 111 ----- 110
Db 6875 ARPDARAQTLNLVEAHDYLEFNLTLEPTQGVKTKATYQSSLLPLQHWKTLQOLD 6934
Qy 111 ----- 110
Db 6935 LVKIVARPETHMNEISDQLPISLVSVANSEPSQFVYKAGLSLVENHALNNSGGLALVF 6994
Qy 111 ----- 110

```
Db 6995 AHDIRECTSRMESLTYGELNTRANQLANYLISQAKRDELICVMEKSVSLYLSILAAYK 7054
Qy 111 ----- 110
Db 7055 AGCGYLPVPETPAARIQILAEADVVKFCLTDSMAPVIADVSRCHIMNVDTTDCSAQSC 7114
Qy 111 ----- 110
Db 7115 TGPQLDFKPTDIAYAVFTSGTTGPKGVLTQBNILSNLEVLSKIYVPEGSRLLOACNQ 7174
Qy 111 ----- 110
Db 7175 AFDVSVFEIFTWTGMCCLCSKDVWFRDFEKAINELEITHLSLPTVAALTDPAHPR 7234
Qy 111 ----- 110
Db 7235 VKFLVTAGEAVTHVHGANAGKGLYQGYGPSETTICTVNSAVESDHVINNICPAFENTS 7294
Qy 111 ----- 110
Db 7295 AFVLTQDDDFQLVPLGGLGELCFGGQVFRGYQNMPELTESKIINHPNYGRIVRSGLGR 7354
Qy 111 -LPDG-----NRVRG----- 120
|||
|||
Db 7355 LLPDGTILIQRTDQKIRGORIELGEISGCLLOFPFVONCAIEVIKTADKERLMFMI 7414
Qy 121 ----- 120
Db 7415 PSGYSKDSYILOPKDNLEBIIKSIYAHADNLPAYMVPDLPVSAIPOTSOGKIDKR 7474
Qy 121 ----- 120
Db 7475 LASDGSALTVEDLNAYSRGADDETSLSATEQOLASALADTLOMSQTSIGRSTSFALG 7534
Qy 121 -VTQ-----LOG----- 126
|||
|||
Db 7535 LDSVSAIRLATNLKRGYSDVSQLRPTIARLAPLLGGESSKOTNEPVTADCEAAVG 7594
Qy 127 ----- 126
Db 7595 SYLHDSVVSQLEHGGTVSQVLPCTPQOAMLSARDTSGSSAYRNKTLFSLHGSVDKX 7654
Qy 127 AC----- 128
|||
Db 7655 ACHEVMQRLHDILRTIFLSTEDSRFPFVQAVLSQWTLPMQECDDIPQLSTLLDSAKAG 7714
Qy 129 -SPTW----- 132
|||
Db 7715 DSIVDHSPPWKIQYRSESTVYLLDMHHALYDANAMSNLLYEVEQLYKQSLSPVSFX 7774
Qy 133 ----- 132
Db 7775 PFLNFMISTVEADALFRDQREFVPKPKRTDVKSGFGTITGRNLNYSPPMVETFLSKH 7834
Qy 133 ----- 132
Db 7835 STTMLSLTOAMMKTLAASQSDVCCGVNVSGRSPVDGIESLVAPCPNTIPVRVDLSK 7894
Qy 133 ----- 132
Db 7895 HRSNLGLVKALQRVNIDSLPYQLTPLRRIOAQAGTNGKRLFDSILVLLQDQDTLLDSAIR 7954
Qy 133 -----SCL----- 135
|||
Db 7955 LEGESGVNDPCIVELAPTNESTYLSLHFNRSYLDDVSNLHOACLSAFASCIRYPSDD 8014
Qy 136 ----- 135
Db 8015 VSDFIDFDADLVAGVLKPDTKMQPVEAAKTRSEKSGSSGDESWSPLELQIRAAYSAVS 8074
Qy 136 ----- 135
Db 8075 SAPEDIRRDTTIYKLGUDSISAIQANRLRKDGLLVQADVMESPSCSELASAVQSRSQ 8134
```

```
Qy 136 --ITEDTQFD----- 143
|||
Db 8135 TPVLDERGFDFEGFKHYRGAALQSHRIATEKVASVRECTPLQSGMLSEYTHSDGHQYFN 8194
Qy 144 ----- 143
Db 8195 HTFYAIEADIDSSKLOSAMSKVLEQHELLRTGTVGTTDDHEHPFVMLTYTEPFDVIDLEQA 8254
Qy 144 ----- 143
Db 8255 SSREGSVVEYSEQKASESVKNNLHLPWRRWSLLGVGRCQLQFSAHHAIFDAESURLIMT 8314
Qy 144 ----- 143
Db 8315 DLQSALSNGYVPTRLTIDGALGHILNSQADVESORTFWSQKLSGAPVTRFPNMTFVRIS 8374
Qy 144 ----- 143
Db 8375 DTEANVELVLNKRKSLBARCOBLGVSMSQVGOANARLLSAYTGESQVTFGVVLSGRT 8434
Qy 144 ----- 143
Db 8435 SPATADAAPCITTLPVSTNTAVDDSQFLKOLMSYNATIOKHQFTPLTNIRNYAESTSEA 8494
Qy 144 ----- 143
Db 8495 LFDLSFVYQRPMDVVDVSSWKIIREKASVELAVSMEALSEDGLGURLTIDPAQVPE 8554
Qy 144 ----- 143
Db 8555 QGKIMLQOMVMIAGLLKPEDAINTSVMSIIPKDPITATDPKYLHEMTEASVKSYSORI 8614
Qy 144 ----- 143
Db 8615 AMEFVDALEDGOISRRHTYRQLDDEEANKIAHLLIDRGVKPGDIIATSFCKPEASFAFY 8674
Qy 144 ----- 143
Db 8675 GILKAGCAFDPTAPAARKAFILEDNSARVLLTSDSIRSELRELTOCDDIILINFENK 8734
Qy 144 ----- 143
Db 8735 NELSTSSVPVSGLPSSVYLYTSGTTGPKCEITHDNVOLVMSFKRLPKGRWTDES 8794
Qy 144 ----- 143
Db 8795 RMLQFASVHFDVSVLEQFWTWIGMRLVCAPRDLILEDIAGFLDTMQIITHLDLTPLSLRL 8854
Qy 144 ----- 143
Db 8855 LDPALVPSLHKGVFITGSGSLKQDQINTMGDVCLFNFGYPTTECTIGTVTFPCVPKEGKP 8914
Qy 144 ----- 143
Db 8915 SNIGMOFNVGCYVLAPGTQTPVLRGAIGELCISGKLVGKYNRPBELTADCFPYLDARF 8974
Qy 144 ----- 143
Db 8975 ERYVRTGDLVRLPHDGSIDFLGRKDNQVKGRLRGORLEIDIEIAVIKRCODIQDTVCIVAKH 9034
Qy 144 ----- 143
Db 9035 PKQDKQLIAFIGINESRKQKPELCPAESTRHLIQTARAACEERLPYVMVPTHFLPIOR 9094
Qy 144 ----- 143
Db 9095 IPLSVNNKVEKLLRQLYADLPTTVIQTATQADSOOSLSDEQKVAQALAEKLLKDDND 9154
Qy 144 ----- 143
Db 9155 LTPSSNIFSLGLSSISAIQFSKKLKASGFTTVQVATVKNPTISRLTKALATSTGRSGGE 9214
```

Qy 144 ----- 143
Db 9215 IADAKOVISACRQRHMGTVTRVLRCKADDIEATAPCTPFOGIIISRLASESSLYFNFSK 9274
Qy 144 ----- 143
Db 9275 FNAQGVLDQKLEGAFNAQALERTQILRTFFIETDDGYVQAVRKTCGLPMTLEVVYDLASVD 9334
Qy 144 ----- 143
Db 9335 GVPFARQKQKRSYNTSHLTVPFEIVVRSGSETEFVSDLHLYDGNSTFIDLMNNVSKLY 9394
Qy 144 ----- 143
Db 9395 NSQADFGKPFVDCIAGFPLRNQGAQKFWLDHLPDVKSAMPPLIDNPASHDVLCTASL 9454
Qy 144 ----- LGVTI 148
Db 9455 DILNQADELRRSLGTV 9471

RESULT 4
T30225 polyketide synthase - Streptomyces hygroscopicus
C:Species: Streptomyces hygroscopicus
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C:Accession: T30225
R:Apelicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A:Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus
A:Reference number: 220782; MUID:96186896; PMID:8635756
A:Accession: T30225
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-10223 <APA>
A:Cross-references: EMBL:X86780; NID:g987088; PID:g987099; PIDN:CAA60459.1
C:Genetics:
A:Gene: rapB
C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protein
C:Keywords: carrier protein
F:54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F:1533-1604/Domain: acyl carrier protein homology <ACPI>
F:1647-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F:2135-2409/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F:3172-3243/Domain: acyl carrier protein homology <ACP2>
F:3287-3681/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F:3778-4052/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F:5160-5231/Domain: acyl carrier protein homology <ACP3>
F:5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F:6776-6847/Domain: acyl carrier protein homology <ACP4>
F:6891-7285/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
F:8411-8482/Domain: acyl carrier protein homology <ACP5>
F:8526-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
F:9012-9285/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F:10069-10140/Domain: acyl carrier protein homology <ACP>

Query Match 30.8%; Score 238; DB 2; Length 10223;
Best Local Similarity 1.0%; Pred. NO. 92;
Matches 104; Conservative 17; Mismatches 25; Indels 9871; Gaps 32;

Qy 1 AAG----- 3
Db 86 AAGKSYCVGGFLDSAGGDFASFFGISPREALMDPQRLVLEASWEAFERAGIEFGSLR 145
Qy 4 ----- 3
Db 146 GSDTGVMFCAYPGGYCVGADLGGFGATAGAVSVLSGRVSYFFGLEGPVTVDTACSSSLV 205
Qy 4 ----- 3
Db 206 ALHQAAVALRQGECSLALVGGVTVMATPOSFVEFSRQRGLASDGRCKAFADSADGTGNAE 265
Qy 4 --GILHLELL-----VAV-----GP----- 16

Db 266 GAGVLLVERLSDAQKRGHQLAVVRGSAVNQDGASNGLSAPNGPSQORVIRAALSNAGLS 325
Qy 17 ----DVFOAH----- 22
Db 326 TAEVDVVEAHGTGTTGLDPIEAQALLATYGDREQLLLGSKVSNLCHTQAAAGVSGVIK 385
Qy 23 ----- 22
Db 386 MVMALQHLVPRTLHVDEPSRRHVDWDGAVALVTENQPMPCRRPRAGVSSFGISGTNA 445
Qy 23 ----- 22
Db 446 HVILESAPPTQAVDDVPPEAEAPVASELVPLVISARTLPALVEYEDRLRAYLAASPGVDV 505
Qy 23 ----- 22
Db 506 RGVASTLAVTRSVFEHRAVLLGDDTVTGTVDPRVVFVPPGQSORAGHGEELAAFPV 565
Qy 23 ----- 22
Db 566 FARIHQVWGLLDVPLEVNETGYAQPALFALQVALFGLLESWGRVDPDAVVGHVSGELAA 625
Qy 23 ----- QEDTERYVL----- 31
Db 626 GYVSGLSLEDACITLV SARARLMQALPPGGMVAVPVSEDEARAVLGEVGEVIAAVNGPSS 685
Qy 32 ----- 31
Db 686 VILSGDETAVLQAAALGKSTRLATSHAFHARMPEMEEFRVTAERLTYTPTPLAMAAG 745
Qy 32 ----- 31
Db 746 DRVTTAEYVVRQVDRVTRFGEQVASYEDAVFIELGADRSRLARLVGVMMLHTDHEAQAAI 805
Qy 32 ----- 31
Db 806 SALAHLVNGVTVDWTALLGDAPATRVLDLPTVAFQHYRWLEGADRAAGHPLLGPAV 865
Qy 32 ----- 31
Db 866 GLAASGVLFTSQVSRSGDLWLQDOTVLPATAPAEMLAAADETGCGLVEELTVEALLLL 925
Qy 32 ----- 31
Db 926 PDGAVQVQVWSEPDAGRRQLSIHARYSDDELWTRLATATLATTGTRVSGWQAGEAWP 985
Qy 32 ----- 31
Db 986 PAGAVPVEARVPSLRGVWRRGSEVFAEVALDDTHDTRFLLHPGLMTAARMAVGEETPAA 1045
Qy 32 ----- 31
Db 1046 WQGLTHARNAELRVRLTSODDETLSVEAVDSTGLPVLARSALDAVPVNEPATND 1105
Qy 32 ----- TNLNICA----- 38
Db 1106 LLTLTWAGIATROOQTGLTGAFEDLAADGAPVDPVVVFTALPGNDDPLAQTRILTAQV 1165
Qy 39 ----- 38
Db 1166 LRTVOEWIGGERFSDSTLVVRVTGTGLAAAVVSGLMRSAQSEHPGFILVESDDDDALTPDQ 1225
Qy 39 ----- 38
Db 1226 LAATVGLDEPLRLISDGRYEVPRLTRAHADPEPERAWDPDGTVLTITGSGVLAGIVARH 1285
Qy 39 ----- 38
Db 1286 LVAERGVRHLLLSRGTPDRALLSELAEALGAAVDTACDVSDRAELARVLARVSPEHPLT 1345
Qy 39 ----- 38

Db 1346 AVIHTAGVDDGVVESLSAORLETVFRPKADGAWHLHELTRDADLAAAFVMSAAAGVMAG 1405
QY 39 ----- 38
Db 1406 AGOONYAANAFDLALAEERRAEGLPALAVAWGLMEDASGLTAQLTDTDRIRRGGLRA 1465
QY 39 ----- 38
Db 1466 ISAEHGMRLPDNASRHSEPLVAAPMEPRDAEVPALLRSLHRPNVRRRAALAGGAQWLAA 1525
QY 39 ----- 38
Db 1526 LAPEERAKALLKVVRDTAATVILGHADARTIPVTGAFRDLGIDSLTAVELRNLAKVTGLR 1585
QY 39 ----- 38
Db 1586 LPATLVFDYPTPAVLAARLGEFTGENPVLVRTASVVGQDEPLAIVGMACRLPGGVSSPE 1645
QY 39 ----- 38
Db 1646 DLWRLVESGTDALSGFPADRGWDAESLFPDPDAVGSKYCEGGFLDSAAASFDAGFFGIS 1705
QY 39 ----- 38
Db 1706 PREALAMPPOORLIMEVWEAFERAGIEPGSVRGSDTGVFMGAYAGGYGAGADLGGFAAT 1765
QY 39 ----- 38
Db 1766 ASATSVLSGRVSFFGLEGPATVDTACSSSLVALYQAGYALRQGECSLALVGGVTVMAT 1825
QY 39 ----- 38
Db 1826 PQSFVEFSRKSGLSDGCRKAFADSADGTGWAEGVGVLLVERLSDAQAGHQLAVVRSS 1885
QY 39 ----- 38
Db 1886 AVNQDGASGLSAPNGPSQGVIRQALANAGLTTAEVDVVVEAHGTGTLGDIPEAQAVIA 1945
QY 39 ----- 38
Db 1946 TYGODRERPLLLGSLKSNIGHAQAAGSVGVIKVMALQHNTVPTLHVDEPSRHVDWAA 2005
QY 39 ----- 38
Db 2006 GAVELVRENQWPQDTPRBRAGVSSFGVSGTNAHVILESAPPAQABEAQVETPVVASD 2065
QY 39 ----- 38
Db 2066 VLPLVISAKTQPALTEHEDRLRAYLAASPGVDTRAVASTLAVTRSVFEHRAVLLGDDTVT 2125
QY 39 ----- 38
Db 2126 GTAVSDPRVVVFPQGWQWLGMSALRDSIVFAERMAECAPALREFVDWDLFTVLDDP 2185
QY 39 ----- 38
Db 2186 AVVDRVDVQPASRMWVSLAAVWQAAGVRPDAVIGHSQGEIAAACVAGAVSMRDAARIV 2245
QY 39 ----- 38
Db 2246 TLRSEAIARGLAGRGAMASVALPAQDVELVDGAWIAAHNGPASTVIAGTPEAVDHLTAH 2305
QY 39 ----- 43
Db 2306 EARGVRVRITVDYASHTPHVELIRDELDTITSSSSQAPVVPWMLSTVDGSWDSPLDVE 2365
QY 44 ----- 43
Db 2366 WYRNLRBEPVGFPAVQLOAEGDTVFEVVSASPVLLQAMDDVVTVATLRRDDGDATRM 2425
QY 44 ----- 47
Db 2426 LTALAAQAVVHGTVDMPAILGTATTATRLDLPYAFQHORYWLRSDVRAAADGHPLLGTGV 2485

QY 48 ----- 47
Db 2486 DLPASDGVVLTGRVSLATHTWLADHAVRGSVLLPGTAFVDLVVRAADEVESDVVDLVBIE 2545
QY 48 ----- 47
Db 2546 TPLLLLPOTGGVQLSVSVGGADESGHRAVMVFSQADNTDTWTRHTVATVSTSDTVSLPEF 2605
QY 48 ----- AQ 49
Db 2606 ASWPPAQAARPVSVADFYDRLAAAAGTEYGPAGFQLOAHRDGTVYAEVVLABEQAOEAAR 2665
QY 50 FRVH----- 53
Db 2666 FAVHPALLDAAMHASVLHTPDTDOOSVRMPFSNNHVQIRATDTAMLURVAATPTTDGWSVR 2725
QY 54 ----- 53
Db 2726 VADDTGRPVATTIGSLVTRPVTTADTLGSAADDLLTLVWTEIPTPOQSSLSVGRYEDLADGD 2785
QY 54 ----- 53
Db 2786 VPPVDVAVYTARPEADGSPDLVQTRTLTAQVLQTAQAMLGGERFTDSTLVVVRTGTGLAA 2845
QY 54 ----- 53
Db 2846 AAVSGLMRSAQSEHPGRFVLVESDDDSLTLHQLAATVGLDEPRLRVNDGRFEVPRLARVN 2905
QY 54 ----- 53
Db 2906 AAEPSETVNDPHGTVLTITGSGVLAGALARHLVTERGVRRHLLLSRTTAD EGLLNELGE 2965
QY 54 ----- 53
Db 2966 LGARVETADCDVSDRAGLARVLAVGVSPEHPLTAVIQTAGALDDGVLETLTAQRDLTVLRP 3025
QY 54 ----- 53
Db 3026 KADGAWHLHELTRNTGLAAAFVMSYSAAGVMPGOGNLAATAFLDALADQRRAEGLPAL 3085
QY 54 ----- 53
Db 3086 ALAWSSETSDLIGURTISABRGMRLFDSASHRGEPLLMAASLDPARAAEVPAALLRSLR 3145
QY 54 ----- 53
Db 3146 RPVARRAASADGGVQWLAALAPAEREKALLKVVCDSAAVVILGHADARTIPVTGAFKDLGV 3205
QY 54 ----- 53
Db 3206 DSLTAVELRNSLVKATGLRLPATMVFDPYPTTALAARLDELFTGENPAPVREPVPAAVD 3265
QY 54 ----- 53
Db 3266 EPLAIVGMACRUPGGVSSPEDLWRLVESGTDVSGPPTDRGWDVEGLFDPDPDAACKSYR 3325
QY 54 ----- 53
Db 3326 AEGGFLLDTAAGFDAGFFGISPREALAMDQQRLLLLEVSWEAFERAGIEPGSVRGSDTVGF 3385
QY 54 ----- 53
Db 3386 IGAFPPVYGAGAAREGYGATAAPNVLSGRLSYFFGLEGPATIMDTACSSSLVALHLLAAQA 3445
QY 54 ----- 53
Db 3446 LRNGCSMALAGGVYMATPEVTFEPARQGLASDGRCKAFADSADGAGFSEGAGLLLVE 3505
QY 54 ----- 53
Db 3506 RUSDARRNGHQLAVVRGSAVNODGASNGFTAPNGPAQORVIRQALANAGLTTAEVDVVE 3565

QY 54 -----LVKWI----- 59
Db 3566 AHGTGTTLGDPIEAQAVIATYQDREQPLLLGLTKSNVGHQTQAAAGVSGVIMWALQHS 3625
QY 60 ----- 59
Db 3626 TVPRTLHVNEPSRHYDWSAGAVELTENQSMPTVGRPRRAGVSAGVSGTNAHVLESAP 3685
QY 60 -----LPE----- 64
Db 3686 PAQSNNAPVATPVVASSELVPLVISAKTLPALTEHEDRLRAYLAASPCADMRAVGSTLA 3745
QY 65 ----- 64
Db 3746 LTRSVFEHRAVLGHDTVTCTGTAVSNPRVVFVPCGQWGLGMSALRGSSVVFPAER 3805
QY 65 ----- 64
Db 3806 MAECAALSEFVDWDLFAVLDDPAVVDRVDVVPQASWAMVMSLAAMQADGVRPDAVIGH 3865
QY 65 ----- 64
Db 3866 SOGEIAAACVAGAVSLRDAARSVTLRQAIARGLAGAMASVALPAHEIELVDGAMTAA 3925
QY 65 -----GAP-----NITANLTS 75
Db 3926 HNGPASTVVAGAPEAVDRVLAVHEARGVRRIANDYASHTPHVELIRDELDDITAGIGS 3985
QY 76 SL----- 77
Db 3986 QAPVVPMLSTVDGTWVEGLDVEYMYRNLRPEVPFDSAVQQLRAEGDVTVEVSASPVL 4045
QY 78 ----- 77
Db 4046 QAMDDVVTATLRRDDGATRLMTALAQAQFVEGTVDMWPAILGATTRVPDLPITYAFQH 4105
QY 78 ----- 77
Db 4106 QRFWAEGADRSVAGHPLGLGVAVELPESDGVLLTGRVSLATHANLADHAVRGSVLLPGTG 4165
QY 78 ----- 77
Db 4166 FVELVRAADEVCDVDELVIETPLLLPSSASVHLSVSGEADSGRRGTVFVSADGA 4225
QY 78 ----- 77
Db 4226 DAWTRHVSATIGVGAALSILPELAAMPQAQPVGLGFYDLTGAGYEYGPAGLOQA 4285
QY 78 ----- 77
Db 4286 WRDGDTVFAEVALAEQAEAAAFVHPALLDAALDAGILNTLDNAEQGVRLPFSWNGVQ 4345
QY 78 ----- 77
Db 4346 VRATGSATLRVADNPTSDGWSVRVADDSGRPVATVDSLVTRPVTADTLGSAADLLTVW 4405
QY 78 ----- 77
Db 4406 TEIPTQOTGLSVGRFEDLADGVPVPEVVVCTALPDSENPIALPDPPDLVQTRTLTT 4465
QY 78 ----- 77
Db 4466 QVLQAOVWLAGERFDTSLVVRTGTGLATAGVSGLMRSQAQSEHPGRFVLVECDNLTQ 4525
QY 78 -----LSVC----- 81
Db 4526 QLAATVGLDEPRLRVCDFEVRPLRANPTPESSPLTIPCDRAWLLEQSHSGTLRLDALV 4585
QY 82 ----- 81
Db 4586 PAETAPRLOSQGEVRVDVRAAGLNFDRDLIALGTYPGEAIVGAAGVWLEVGPEVQDLA 4645
QY 82 -----GWSQT----- 86

Db 4646 PGDRVFELVGGFGGFAVAIADRRMLGVIPOWSEFTTAASVPVVFATAYYGLVDLAGLSAGE 4705
QY 87 ----- 86
Db 4706 SVLIHAAAGVGWMAQTQIARHLGARIYATASTGKHVLRAGLEDARIGDSRTTGRPMV 4765
QY 87 ----- 86
Db 4766 LDTTDSRGVDVNLNSLGD FVDASDLLPRGGRFVEMGKTDIRDPHQVTADRPCTSYQAF 4825
QY 87 ----- 86
Db 4826 DLMDAGPDLREIITADLLALFAQGVLLPLVPRAWDIRQAREAFSWSRARHIGKIVLTVP 4885
QY 87 ----- 86
Db 4886 QQLDANGTVLVTGSGVGLAGIAARHLVAEOGVRRHLLLSRSTPDDALINELGELGARVDT 4945
QY 87 -----INPE----- 90
Db 4946 AICDVSDRAGLARILAGVSPHEPLTAVIHTAGALDDGWVESLTAQQLTVLRPKADGAWH 5005
QY 91 ----- 90
Db 5006 LHELTRDADLAAAFVWYSSAAGVLGSGGQGYAANAFDLAEOQRGEGLPALAVAGLW 5065
QY 91 ----- 90
Db 5066 EDASGLTAEMTDDTRDIRRGGLRAISAGHGMGLLDAARHGEPVLLAAAMEPVREAEVP 5125
QY 91 ----- 90
Db 5126 ALLRLHRPVARRAASTGSSVQWMLARLAPVEREKALLKLVCDGAATVLGHADASTIPAT 5185
QY 91 ----- 90
Db 5186 AAFKDLGIDSLTAVELRNSLTAKTGLRPLATLPDPTPTALAAALGEVFWGETPVPVRT 5245
QY 91 ----- 90
Db 5246 SVSVVAQDEPLAIVGMACRULPGGVSSPEDLWRLLSGETDAVSGPTDRGWDVNLFGPAA 5305
QY 91 ----- 90
Db 5306 GDSYRLOGGFLDAAAGPDASFFGISPREALAMDQQRLVLEVSWEAFERAGIEPGSVRGT 5365
QY 91 ----- 90
Db 5366 DTGVFMGAYPGGYGIGADLGGFGATASAVSVLSGRVSFFGLEGPAITVLCACSSSLVAL 5425
QY 91 ----- 90
Db 5426 HQAGYALRQGECSALVGGVTVMATPQTFVEFARQGGLAGDGRSKAFADSADGAGFSEGV 5485
QY 91 ----- 90
Db 5486 GVLLVERLSDAQKHQVLA MLRSSAVNQDCASNGLTAPNGPSQORVIOAALSNAAGLAH 5545
QY 91 ----- 90
Db 5546 EVDVVEAHGTGTTLGDPIEAQALLATYQDREQPLLLGSKVSNLGHQTOAAAGVSGVIKHW 5605
QY 91 ----- 90
Db 5606 MALQRFVPRTLHVDEPSRHVMSAGAVALTENQPHDMPGRARRAGVSSFGISGTNAHV 5665
QY 91 ----- 90
Db 5666 ILESAPPTQPADNAVIERAPEWLPWVISARTQSALTEHEGLRAYLAASPGVDMRAVAST 5725
QY 91 ----- 90

```
Db 5726 LAITRSVFHRAVLLGDDTVTGTAAATDPRVVFVPGGQSQRAGMGEELAAAFVFAIRHQ 5785
Qy 91 ----- 90
Db 5786 QVNDLLDVPDEVNETGYAQPALFALQVALFGLLESWGVPRDAVVHGSVGEELAAAGYVSL 5845
Qy 91 ----- 90
Db 5846 WSLLEDACTLV SARALMQALPAGGVMAVVPVSEDEARAVLGEVETAAVNGPSSVVLSGD 5905
Qy 91 ----- 90
Db 5906 EAAVLOAAEGLGWKTRLATSHAFSARMEPMBEEFRAVAEGLTYRTPQVMAAGDQTTTT 5965
Qy 91 ----- 90
Db 5966 EYVVRQVRDTRFGEQVASYEDAVFVELGADRSLARLVGVAMLHGDHEAQAAVASALAH 6025
Qy 91 ----- 90
Db 6026 YVNGVTVDMPALLGDAPATRVLDLPTVYAFQHRVWLEGTDRMAAGHPHLLGEAVAVPGSD 6085
Qy 91 ----- 90
Db 6086 GVLLTGRVSLATHPWLADHAVRGSVLLPGTGFVELVVRAADEVGCDVIDELVIETPLLP 6145
Qy 91 ----- 90
Db 6146 QTGGVQLSVAVESDDSGRRRAVTVFSRADNVDTWTRHATVTVNSDITLSPDLASWPPA 6205
Qy 91 ----- 90
Db 6206 QAPVNVADFYDQLTVVAYEYGPAPFQGLQAARWGDVTVAEVALAEQAOAARNAVHPA 6265
Qy 91 ----- 90
Db 6266 LLDAALHAGILNTPDTEQGVRLPESMNQVQVHAVGSAMLRVAVTQAADGWSVRVADTGR 6325
Qy 91 ----- 90
Db 6326 PVATIGSLVTRPVTAOTLGSTTDDLLALTWTEIPAPOETSLTIGRPFEDLADADVPEVV 6385
Qy 91 ----- 90
Db 6386 VFTARPDSSENPLAQTRTLTAQVLQKIQAMLABERFTDSTLIVRIGTGLAAAVSGLMRS 6445
Qy 91 ----- DDT ----- 93
Db 6446 QSEHPGRFVLVESDDTLAPDQLAATVELDEPLRLISDGRFVPRLTRTHAAEPESEKV 6505
Qy 94 ----- 93
Db 6506 WDPDGTVLITGGVLAGIAARHLVAERGVRRHLLLSRSAPDGGALISELGELOAVATV 6565
Qy 94 ----- 93
Db 6566 CDVSDRPLVRVLADVSPERPLTAVIHTAGVDDGVVESLTAQRLDVTVLRPKADGAMHLH 6625
Qy 94 ----- 93
Db 6626 ELTRDADLAAFMVYSSAAGVFGSAGQGNAYANAFDLALAEQRRRAEGLPALALAWLWEG 6685
Qy 94 ----- DP ----- 95
Db 6686 TSGTLANTDTHDRIRRSGRMAISAEBGMRLFDGASRRRDPVLVAAMEFVREAEVPAM 6745
Qy 96 ----- GHADL ----- 100
Db 6746 LRSLRPVARRAASGGAVRCLAALAPEERAKALVKVVCDSAAATVLGHADVDSIPTVTAFR 6805
Qy 101 ----- IIII ----- 100
Db 6806 DLGVDSLTAVELRNSLTATGLRLPATLVFDYPTPGALAAARLEBELFTGENPVQVTRPVSA 6865
```

```
Qy 101 ----- 100
Db 6866 VQDEPLAIVGMACRLPGCVSSPEDLWRLVSGTDAISGFTDRGWDVENLFDSDPDAAG 6925
Qy 101 ----- 100
Db 6926 KSYCVEGGFATAANFDASFFGISPREALAMPQORLVLEVSWEPERAGIBPGSVRGSD 6985
Qy 101 ----- 100
Db 6986 TGVFMCAFFGGYGIGADLEGYATAGLNVLSGRLSYFFGLEGPATVTDTCSSSLVALHQ 7045
Qy 101 ----- V 101
Db 7046 AGVALRQGECSLALIGGVTVMATPHTFVBFSRQGLASGRCKAFADSADGTGWSGVG 7105
Qy 102 LYITRF ----- 107
Db 7106 LLVERLSDAQAKGHQVLAIVRSSAVNQDGSNGLSAPNPSOORVIRQALANAGLTAEV 7165
Qy 108 ----- 107
Db 7166 DVVEAHGTGTTLODPIEAQVIATYQDRDQPVLLGSVKSINVCHTQAAAAGVSGVIKVMMA 7225
Qy 108 ----- 107
Db 7226 LQHGLVPRTLHVDEPSRHVDWDGAVELVTENQSWPEAGRPRRAGVSFGVSGTNAHVIL 7285
Qy 108 ----- 107
Db 7286 ESAPPTQAVDDVRPADAPVVASVMASELPLVISAKTOSALAEYEGRLRAYLAASPGVDM 7345
Qy 108 ----- 107
Db 7346 RAVASTLAMTRSVFHEHRAVIVGDDTVSGTAATDPRVVFVPGGQSQRAGMGEELAAAFV 7405
Qy 108 ----- DLE ----- 110
Db 7406 FARIHQVWDLLDVDPDEVNETGYAQPALFALQVALFGLLESWGVPRDAVIGHSVGELAA 7465
Qy 111 ----- LPDGN ----- 115
Db 7466 AYVSGLWLSLEDACTLV SARALMQALPAGGVMAVVPVSEDEARAVLGEVETAAVNGPSS 7525
Qy 116 ----- 115
Db 7526 VVLSGDEAAVLOAAEGLGWKTRLATSHAFSARMEPMBEEFRAVAOGLTYHAPGVVMAAG 7585
Qy 116 ----- ROVR ----- 119
Db 7586 DRVMTAEYVVRQVRDTRFGEQVASYEDAVFVELGADRSLARLVGVAMLHGDHETQAAI 7645
Qy 120 ----- GVT ----- 122
Db 7646 GALAHLVYNGVTVDMTALLGDVPVTRVLDLPTVYAFQORYMAEVRGSADVSGAGLDAVGH 7705
Qy 123 ----- 122
Db 7706 PLLGAVVALPGSDGVLTLTGRVSLATHAWLADHAVRGSVLLPGTGFVELVVRAADEVGCDV 7765
Qy 123 ----- 122
Db 7766 VDELIVEAPLLLPTQGVQVSVSVEADESGHRVVTIFSRADNADTWVRHRSVAVISTDS 7825
Qy 123 ----- 122
Db 7826 TVSLSEFASWPPAQAPQVDTGTGYDPLMAEGYEGPAGFOGLOAWRYGDTVFAEVALAE 7885
Qy 123 ----- 122
Db 7886 QVREAAARYPSVHPALLDAAALHACTLECAEAEVGVGLPFFSWDGVVRHAGGSAMLRVATQA 7945
```


Qy 123 ----- 122
 Db 7946 ADGWSVRVADEIGRPVANVGLVTRPVPTADALGSAADLLALTWAEIPAPEGTLTVGRF 8005
 Qy 123 ----- 122
 Db 8006 BELVSDGDPVPEVAVFTALPDNDPPLEQTRKLTGQVLQAVQEWLGGERSFSDTLVVRT 8065
 Qy 123 ----- 122
 Db 8066 GTGLAAAGVGLMRSQAQSEHPGRFVLVESDDDALADQLAAAVGLDEPRLRISDGRFEAP 8125
 Qy 123 ----- 122
 Db 8126 RLTRTHAAPESEKVPDGTVLJTGSGVLGAGIARHLVTERGVRHLLLSRSAPDEAL 8185
 Qy 123 ----- 125
 Db 8186 INQLGELGARVETAACDVSDRAALAQVLGVSPEHPLTAVIHTAGVLDGCVVESLTAQRL 8245
 Qy 126 ----- 125
 Db 8246 DAVLRPKADGAWNLHELTRDADLAAAFVMYSSAAGVLGSGGQGNAAANAFDALAEORHA 8305
 Qy 126 ----- 125
 Db 8306 EGLPALAVAWGLWEDASGLTAQLTDTDRIRRGGLRAISAHEGMGLFDSASRHSEPVIV 8365
 Qy 126 ----- 125
 Db 8366 AAPMEPVRDAEVPALLRSLRHPVARRAAAGGARWLALAPAEREKALLKLVSDCAATVL 8425
 Qy 126 ----- 125
 Db 8426 GHADTSTIPATFAFKDLGIDSLTAVELRNSLAKATELRLPATLFDYPTPTALAAARLDEL 8485
 Qy 126 ----- 125
 Db 8486 FTGENPAPVRGPVSAVAODEPLAIVGMACRLPGGVSSPEDLWRLVESGTDIAISGPTDRG 8545
 Qy 126 ----- 125
 Db 8546 WDVENLYDPDPAPGKSYSQGFLDAAAGFDASFFGISPREALAMDPOORLMLVSWEA 8605
 Qy 126 ----- 125
 Db 8606 FERAGIEPGSVRGSDTGVFIGAYPGGYGIGADLGGFTTAGAASVLSGRVSYFFGLEGPA 8665
 Qy 126 ----- 131
 Db 8666 FTVDTACSSSLVALHOAGYALRQGECSLALVGGVTVMPTPQTFVFSRORGLSADGRCKA 8725
 Qy 132 ----- 131
 Db 8726 FADAADCTGAEGVGVLLVVERLSDAQHQHILAVRSSAVNQDGNGLSAPNGPSQR 8785
 Qy 132 ----- 131
 Db 8786 VIRAALSAGLAPHEVDVVEAHGTGTLGDPTEAQAVIATYGGGGEPLLLGLSKSNVGH 8845
 Qy 132 ----- 133
 Db 8846 TQAAAGVSVIKVMALQHSMPRTLHVDEPSRHRVDSAGAVELVAENQFNPETGRPRRA 8905
 Qy 134 ----- 133
 Db 8906 GVSSFGISGTNAHVILESAPASQSGDTAGSTPVLVSELVPLVISAKTQPALTEHEDRLRA 8965
 Qy 134 ----- 133
 Db 8966 YLAASPCVDIRAVASTLAVTRSVFEHRAVLGDETGTGTAVSDPRIVFVFPQGMQWLGM 9025
 Qy 134 ----- 133

Db 9026 GSALROSSVVFERMAECABALSEFVDMDLFAVLDDPAVVDRVDVVQPASWVMVSLAAV 9085
 Qy 134 ----- 135
 Db 9086 WQAAGVRPDAVIGHISOGEIAAACVAGAVSMRDAARIIVTLRSOAIARGLAGRGAMASVALP 9145
 Qy 136 ----- 135
 Db 9146 AQDVELVDGAWIAAHNGPASTVIAGTPEAVDHLVTALRQRGAGNADHVDYASHTPHVELI 9205
 Qy 136 ----- 139
 Db 9206 RDELLDITSSSSQDPLVPHLSTVDGTWVDSPLDGEVWYRNLREPVGPHPAVSOLQAQGD 9265
 Qy 140 ----- 139
 Db 9266 TVFVEVASPVLMOAMDDVVTVATLRRDDGATRMLTALAQAAYVHGVTVDMRAVLGDVP 9325
 Qy 140 ----- 139
 Db 9326 ATRVLDLPTTAFQHORVWAEAGRSADVSAAGLDAVGHPLLGAVLAMPGSDGVMLTGRVSL 9385
 Qy 140 ----- 142
 Db 9386 ATHAWLADHAVRGSVLLPGTGFVELVVRAADEVACDVVDELIVEAPLLLPQTGGVQLSVS 9445
 Qy 143 ----- 142
 Db 9446 VGEADESGHRAVTVFSRADSADAMVRHVSATVSVSDTTVTSTLTAMPPQAOKPVDVAGF 9505
 Qy 143 ----- 142
 Db 9506 YDQLTRAGYEGPAGFQGLQAARDDGTVAEVALAEQOTQDAARFAVHPAVLDAALHAGI 9565
 Qy 143 ----- 142
 Db 9566 LNTPDADRDTVRLPFSMNHVQVHVTSATLRVAMTRVADGWGVRVADDIGRPVATIGSLV 9625
 Qy 143 ----- 142
 Db 9626 TRPVAADALGSAVDDLFALTWTEIPVSQQGVTVGKFEDLADGEVPMPPVVVFTALPDSG 9685
 Qy 143 ----- 142
 Db 9686 DPLAQTRRLTAELVLOAVQVWLAGERTFTDSTLVVVTGTGLAAAAVSGLMRSAQSEHPGRFV 9745
 Qy 143 ----- 142
 Db 9746 LVESDDDDTLTPDQLAATVGLDEPRLRVIDGRYEAPRLTRTGTVAEPEPEGVMDPDGTVLIT 9805
 Qy 143 ----- 142
 Db 9806 GSGVLGAGIAARHLVAERGVHRHLLLSRSAPDEALISELAEAGAAVDTAVCDVSDRAGL 9865
 Qy 143 ----- 142
 Db 9866 ARVLAVGSPDHPLTAVIHTAGVLDGCVVESLTARRLOTVLRPKADGAWNLHELTRDIDLA 9925
 Qy 143 ----- 142
 Db 9926 AFVMYSSAAGVLSAGQGNAYANAFVDALAEORRAEGLPALALAWGLWEDASGLTAKLT 9985
 Qy 143 ----- 142
 Db 9986 GTDHDRIIRSGLTITTAERGMLFDIASROGEPVLVATPMEPVEVEVPALLRLLLRPVA 10045
 Qy 143 ----- 146
 Db 10046 RRAASTGDSNAQWLVLGLAPEERAKALLKVVTRDSATVGLHADARSIPATGAFKDLGV 10102
 RESULT 5

T14593
syringomycin synthetase - Pseudomonas syringae pv. syringae
C:Species: Pseudomonas syringae pv. syringae
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 01-Dec-2000
C:Accession: T14593
R:Guenzi, E.; Galli, G.; Grgurina, I.; Gross, D.C.; Grandi, G.
J. Biol. Chem. 273, 32857-32863, 1998
A:Title: Characterization of the syringomycin synthetase gene cluster. A link between p
A:Reference number: Z18153; MUID:99047670; PMID:9830033
A:Accession: T14593
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-9376 <GUE>
A:Cross-references: EMBL:AF047828; NID:G3510628; PTD:G3510629; PIDN:AAC80285.1
C:Genetics:
A:Gene: syrE
C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:441-896/Domain: acetate-CoA ligase homology <ACL1>
F:914-981/Domain: acyl carrier protein homology <ACP1>
F:1529-1984/Domain: acetate-CoA ligase homology <ACL2>
F:2002-2089/Domain: acyl carrier protein homology <ACP2>
F:2613-3071/Domain: acetate-CoA ligase homology <ACL3>
F:3089-3156/Domain: acyl carrier protein homology <ACP3>
F:3700-4158/Domain: acetate-CoA ligase homology <ACL4>
F:4176-4244/Domain: acyl carrier protein homology <ACP4>
F:4768-5224/Domain: acetate-CoA ligase homology <ACL5>
F:5242-5310/Domain: acyl carrier protein homology <ACP5>
F:5834-6280/Domain: acetate-CoA ligase homology <ACL6>
F:6298-6366/Domain: acyl carrier protein homology <ACP6>
F:6892-7352/Domain: acetate-CoA ligase homology <ACL7>
F:7370-7437/Domain: acyl carrier protein homology <ACP7>
F:7990-8440/Domain: acetate-CoA ligase homology <ACL8>
F:8458-8526/Domain: acyl carrier protein homology <ACP8>
F:9015-9083/Domain: acyl carrier protein homology <ACP9>
F:946, 2034, 3121, 4208, 5274, 6330/Binding site: phosphopantetheine (Ser) (covalent) #status

Query Match 30.7%; Score 237; DB 2; Length 9376;
Best Local Similarity 1.2%; Pred. No. 76;
Matches 94; Conservative 24; Mismatches 28; Indels 7809; Gaps 29;

QY	1	AAGG-----	4
DB	980	AAGGSEVVVPA NI PEHC DR ITPQ ML PL IN LSQAQIDHVV KM PGGVANVQDIYPLAPL	1039
QY	5	-----	4
DB	1040	OAGILYHHISAEQGD P YTKALFALSDRAQLD P FGALQGVINRH D ILRTAVLWEDL D EP	1099
QY	5	-----	4
DB	1100	VQVVL RQ ANLHVTELL LD PDVGPVDEQLHQ R FRDRH Y LDV R TAPLMRIV F SHDPV N DRW	1159
QY	5	-----	4
DB	1160	LAMLLCHHLVSDATSLSVILQEIQAHLIGOGNALGEAVPYRNVAQ R LVSEAQEAFF	1219
QY	5	-----ILHL--	8
DB	1220	RDMLGDI D EPTLPFLQDVQ D GRNLEASVTLAAELNRLRAQARQAGVSASLMHLAW	1279
QY	9	-----	8
DB	1280	ARVLGNVSACQVV F GTVLLGRMQGADRALGMFINTLPLRV D IGATTVAEGLKATHE	1339
QY	9	-----	8
DB	1340	RLTALLGHEAPLVLAQRCSGVAAPTPLFSALLN Y HSVSDVRQKLPGLGSSLRHRRP	1399
QY	9	-----	8
DB	1400	YQLPVDNLVDLGDAFTLTIAQVOQISATRIGEYMQVALRNLDALHTPQAALNSLSIL	1459

QY	9	-----	8
DB	1460	PDDE R ELLTGFNDTAHPY P RDVLIHOLIEQQAORPDACA V RGD S GTLLTYAELNOQAN	1519
QY	9	-----ELLVA-- 	13
DB	1520	QLA H RIELGVEPD T RVAVSLRGAEMVVALLGILKAGAYV P IDPDL P ARQAYM L SDS	1579
QY	14	-----	13
DB	1580	APRAVLT S HELLADLPDLGVPALVLDGRD S SALLKKQPTGNPD K ALDLPQHLYAVLYT	1639
QY	14	-----	13
DB	1640	SGSTGTPKGYMNEHLGVNRLWARDAYOVNSQDRVLQKT P FGFDVSVMEFFLPLLTGAE	1699
QY	14	-----	13
DB	1700	LVNARPSGHQDPYLAQVISDAGITLLH F VP S MLDV F LEH S TRDPQLRRVLC S GEALP	1759
QY	14	-----	13
DB	1760	RALQ R FEQHLKGV L HNLYGPTAAIDVTAWECRPTD P CGSV P ICRPIANI Q MHVLDAL	1819
QY	14	-----	13
DB	1820	GQLQPMGVAGELHIGGIVAGYLNQ P LSAERFIADPF S NDPQARLYKTGDVGRWLANG	1879
QY	14	-----	13
DB	1880	ALEYLGRND F QVKIRGLRIEIGIEAALAKHPA V EA V TAREDI P GD K R L VAYYTQTAE	1939
QY	14	-----	13
DB	1940	HTAVDLET L RGH L QQLVPEYMWPAIYVLL E AMPLT S NGKLD R KALPAPDGDALISRGYEA	1999
QY	14	-----	13
DB	2000	POGEIEQIAVIMQDLLGVEQGRHDH F ELG H SL L AVSLIDRLRKODLN V NTVFTA	2059
QY	14	-----	13
DB	2060	PSVREMA L AI S OEKQALFRAPANRI P AHCTQLT P DM L PLV S VAQIELIASAVPGGAN	2119
QY	14	-----	13
DB	2120	IQHIYPLAPLOEGILFHYLLNLERDAYLYRSTIE F DSRARLDA F LEGLQTV I DRHDV L RS	2179
QY	14	-----	13
DB	2180	SVHWGLPQAVQVVHQAQSP I HTLT R PDE D ALSQ L DR L SDPCLRL R DLRQAPLLLAYI	2239
QY	14	-----	13
DB	2240	ARPD S ERWLLALIDHHMISDHVTLEIL E IRLMRGSAELLPPOPYREFVAQT L ASP	2299
QY	14	-----	13
DB	2300	SSAHEY F TGR L ADVDSPTAFELLEVOG D NDVESELS S DL C AR I RTQAR G CMSP	2359
QY	14	-----	13
DB	2360	AVLFHVAQV L ARCTGRDDV F GTAVTGRLOGT L GAERAMGFMNTLPVRVQLATQSVQ	2419
QY	14	-----	13
DB	2420	ELVNATH R DLSELLSH E QASLAAQRCSSVATGV P LFSSLLN Y RH N EDSOLQMPGLRL	2479
QY	14	-----	13
DB	2480	DSABRTN P LC S VNDYGS D LGILLIH S VQPADPQRLCMMQCALEQLT A LHTPQKEVT	2539
QY	14	-----	13

Db 2540 QLDVLPAAERNLLLETFNCTRODYPTDLCIQLHLEFAQVQTPDPAIAVAFAQRLSYAELN 2599
Qy 14 -----VGPD----- 17
: |||
Db 2600 ROANRLAHLILGIGCPDPRVAICVERGVMVGLLVKAGAAVPLDPAYPAERLAYM 2659
Qy 18 ----- 17
Db 2660 IEDSOPALMTORHLOEYLPTLTPLVLDDDDQRTFTTERDDNPVVEALGVNRLAYVIYT 2719
Qy 18 ----- 17
Db 2720 SGSTGNPKGMIEHRGLVNYSDAARLFDLSPTDVTLOQNTLNFDSVEEIPALLAGAT 2779
Qy 18 ----- 17
Db 2780 LTPSREIFGSEGTENHGINPTVLHTAAHHTLVAEWHKQPOVAEORLOHVLINVTGDA 2839
Qy 18 ----- 17
Db 2840 LSAQKLMWDEVRPAHTRLINTYGPTEATVCTAAYVSHDAAGSEGSNATIGKPMANT 2899
Qy 18 ---VFOAQOE-----D 25
: |||
Db 2900 RIYLLDAHQOPVPYGVAGEIYIGDGVARGYLNLEEVNAERFLADPFSESPPARMYKTD 2959
Qy 26 TERYV----- 30
: |||
Db 2960 LARYNADGRIEYLRNDFQVKVGRFRIELGEIEARLGNCTGVKEAVVIAREDNPGDKRLV 3019
Qy 31 -----LT----- 32
: |||
Db 3020 AYVVAQOSOLTAADRLAELAPOLAEYMLPSAFVLLDELPLTPNKKLDRKALPAPDADA 3079
Qy 33 -----NLN 35
: |||
Db 3080 ISRGYAPOGETETQIVAIWDLGIEQVGRHDFPELGGHSLLASVLIERLRKQGLNLN 3139
Qy 36 I----- 36
: |||
Db 3140 VKTFTAPTREMALAIQDKHVLFOVPANRIPAHCTQTPDMLPLVELSVAQIELITSA 3199
Qy 37 ----- 36
Db 3200 VAGGAANIQDIYPLAPLODGIILFHYLLNRERDAYLMRSMIEFDSRARLDAFLEGLQTV 3259
Qy 37 ----- 36
Db 3260 RHDILRSSVHNIGLPQAVQVVRQAQLPVHTLTLPTEEDALSQDLRLSDPGRLLDLRQA 3319
Qy 37 ----- 36
Db 3320 PLLAYIARDPNSERWLLALIDHMHSDHVTVELILEIRLLMRGOSADLLPPQVPYDFV 3379
Qy 37 ----- 36
Db 3380 AOTLASPSSEAHEYFTRRLADVDSPTAPPELLEVQDGNDEEAKLALNSDLCIRTOA 3439
Qy 37 -----GAE----- 39
: |||
Db 3440 RERGMSPAVLPHVAQAQVMARCTGRDDVVFGTVVTRGLOGTAGAERAMGMFNTLPVRVQ 3499
Qy 40 ----- 39
Db 3500 LTTQGAQELVMATHRDLSELLSHEQASLALAQRCSSVATGVPLFSSLLNRYHQEDNRIQ 3559
Qy 40 ----- 39
Db 3560 WPGMRLLDGTERTNYPLCLSVNDYGSGLDIIHSMQPANPQRLCAMMOCALEQTLDAHAH 3619
Qy 40 ----- 39

Db 3620 TPQMAVTQDVLPAARNLLLETFNCTRODYPTDLCIQLHLEFAQVQTPDPAIAVAQOGR 3679
Qy 40 ----- 39
Db 3680 LSYADLNQRANRLAHLILSILGIVPDDRVAICVERGVMIGLLGLLVKAGAAVPLDPAYP 3739
Qy 40 -----LLRD-----PSLGAO- 49
: |||
Db 3740 AERLAYMITDSOPAAALLTLPGLQDRLPALSPMLPLVLLDDEQYOGLAECDDNPVWPLGVN 3799
Qy 50 ----- 49
Db 3800 LAYVIYTSSTGNPKGMIEHRGLVNYSDAARLFDLSQSDVTVLQONTLNFDSVEEIPF 3859
Qy 50 ----- 49
Db 3860 ALLAGATLAPSREIFGSEGTETHGIOPTVLHLTTAAHHTLVAEWHNQPAAEORLOHVRL 3919
Qy 50 ----- 49
Db 3920 INVTGDALSQKLMWDEVRPAHTLLINTYGPTEATVCTAAYVSYDAAAGSEGSNATI 3979
Qy 50 ----- 49
Db 3980 GKPMANTRIYLLDAHQOPVPYGVAGEIYIGDGVARGYLNLEEVNAERFLADPFSES 4039
Qy 50 -----FRVHL-----VKHVL----- 60
: |||
Db 4040 RMYKTGLARYNADGRIEYLRNDFQVKVGRFRIELGEIEARLGNCKGVKEAVVIAREDN 4099
Qy 61 ----- 60
Db 4100 PGEKRFVAVVQAQPTQITAAELRAELAPOLAEYMLPSAFVLLDELPLTPNKKFDRKALP 4159
Qy 61 -----TEPEGAPNIT-ANLTSSLLSV----- 80
: |||
Db 4160 APADDAFASREHVEPOGATEIALAQIMQSLDLERVRGRHDFELGGHSLLAMRLISOAR 4219
Qy 81 ----- 80
Db 4220 HOLGVELGALAFAPHEVSAALAVIAOAGRSKFPDIPVARDQAWPLSFGOORLWFLAQM 4279
Qy 81 ----- 80
Db 4280 EGASAAHYHYPAGLSLHGNLNLKALQALERIVARHEGLRTTFMQDDGQPVORISPADTG 4339
Qy 81 ----- 80
Db 4340 FNLQMHDLQGLADAEKQLALASEESLQSPDLOQGLIRGLIRMAEDHHVLLLTLLHHIV 4399
Qy 81 -CGHS----- 84
: |||
Db 4400 SDGMSVDVLTRELSALYAASFQDQDDPLAPLEQYLDYAVWQRWLSGDVLOQOSNFMQO 4459
Qy 85 ----- 84
Db 4460 TLADAPALLMLPTDRPALQDYAGAAALPVVFDKDLTRGLKALSORRGSTLFMTVMAAWA 4519
Qy 85 ----- 84
Db 4520 GLLGRLAGQDDVWIGTVPANRTRSEVGLVGLFVNTLAIKRVDSLDPKPTAETLLARVKOOT 4579
Qy 85 -----QTINP----- 89
: |||
Db 4580 LDAQAHQDLPPEQVVVEVINPLRSLSHSPIFOAMLSWENNEASDLTGLDNTLKSIELA 4639
Qy 90 ----- 89
Db 4640 AQFDLTLDMAEVDQVCTLEVATATLFDSTMRRYLVGYFORVLEAMVADDQVLEHVALL 4699
Qy 90 ----- 89
Db 4700 GADEREHLLAGLNATEAPFPQDRTIHQLFEBRVOAQDAIAVAFGAORLSYAE LNQRANR 4759

```
QY 90 -----EDDT 93
Db 47:0 VAHHLISLGKPPDRVAICVERGEMLIGVLKAGAAVPLDPAYPAERLAYMIEDST 4819
QY 94 -----
Db 4820 PSALLAQRDVQAHPLTLDLPLVLLEDQRTTLSERDDNPVETLGAHNLAYVIYTSGSTG 4879
QY 94 -----
Db 4880 VPKGMVHRGLFAVSAANQOLVALHAPLNLQWAGFFVFSADLIRSLAFGGTLVLCPR 4939
QY 94 ----DP-----GHADLVL-----
Db 4940 ETLMDPPALYRLISEESIGFADFVLAVLNALLGWVEGTGHDLSFMRTVVCGSDIWTASHA 4999
QY 103-----
Db 5000 RQLRKLGDHVQVQVQAYGVTEASIDSTCFEATSQVDGVLPIGRALANTRIYLLDELQ 5059
QY 103-----
Db 5060 AVPPGVAGELYIGGAGIARGYLNLPLQTAERFMDNPFVAGERLYRSGDMARYRADGNIEF 5119
QY 103 ----DP-----YITR-----
Db 5120 LGRNDSQAKLRGLRLELGEIEARLAEVAGVRESLVIREDSGGTPKLIAYFVEYATRDES 5179
QY 107-----
Db 5180 GPALTPRALRQQLQLNPEYMI PAAFVRMAALPLSANGKLDRRALPEPDADAFDQHDFA 5239
QY 107-----
Db 5240 ADGPLETAIAAIWADVILGVAQVRHDDFFALGHSLLVMRLVAQVRQOLNLEVSPTSFFA 5299
QY 107-----
Db 5300 APVLRQFAERLGNQDNARVAIKDVQRSGALPLSYAQRWLFLAQLEGGSAAYHIPAGLR 5359
QY 107-----
Db 5360 LRGNLQASLORALDRIVARHEALRTTFVQEQGQPAEQRIASAEFTGRLOQLVLAGOTDA 5419
QY 107-----
Db 5420 EDKLLAIAAQEASEHFDLVNGPLVRGRLVRMANDNHVLLVTMHHILSDGWSGTGVNREL 5479
QY 107-----
Db 5480 ALYAFRQAGADPLPALPVQVVOYALWQDNLSDGVLOQQRQYWQALAGAPALLTLPTD 5539
QY 107-----
Db 5540 RPRPAQDYSQTLLELVDLTOLTRGLKALSQRHGSLSLFTVMGAWAALLGRLSGQDDWI 5599
QY 107-----
Db 5600 GTPVANRMSAEVENLIGFFVNTLAIRVDLSGTPSVEALVRQVKORTLAAQANODLPFEQV 5659
QY 107 -----FDLE-----
Db 5660 VEVVOPQRSLSHSPIFOAMLSWQNNEDTALVLGDLTLQGVAVAGDTAKFDLALDIGEVDG 5719
QY 111-----
Db 5720 QLIGTLEVATALFDESTMRRYRGYFLRLLEAWAADDQQLVQVPLDPTAEREYLLKDINA 5779
QY 111-----
Db 5780 TERTYPVQGMHRLFEAHAEAAPOAIAVRQGEQTLTVAELDSRANALAQHLRKHGVGPGT 5839
```

```
QY 111 -----
Db 5840 RVAILLDRSVELLASMLATLKGAAYLALDRLAPEERLFMLEDEAIMLSRDLTAPD 5899
QY 111 -----
Db 5900 MTPRLDLDTLELSALNOEPVPVATEVAGETPACIIYTSGSTGVPKGVIVTHNGIVRLVOD 5959
QY 111 -----
Db 5960 NGYDFSAEDRVAFSSNPAPDASTPEIMGALLNGCQSVIIEPQVLLPEVAFALLKRHGV 6019
QY 111 -----
Db 6020 TAMISSTALFNLYAGLIPALAGLRMIMCGGERADPASPFRVREHSAQVRLFNQYGPTEG 6079
QY 111 -----
Db 6080 TTCATSYEIFDVLPTLSLPIGKPNANVRYVLDARREPVMGVVGDIIYIGGTGVALGYL 6139
QY 111 -----LPGNRQ-----VRG-----
Db 6140 NRPELTAERFSEDPFSQAGARLYRTGDLARWLPGDNLEVLARNQGVKVRGPRVELGEI 6199
QY 121-----
Db 6200 ESVLHLYDGRNVSVVAHEASPGDTRLVAYTVHAGVAAAPDPEDLRAQLSADLAEMYWPS 6259
QY 121-----
Db 6260 VFVRDLALPTLNGKVDQKALPVPDLNAMFERTYVAPEGATEQALAEI FOELLGLSERVGR 6319
QY 121-----
Db 6320 HDGFFELGCHSLLAAQLVSRVRQQLNGDMALROLFNHPHTVAELAKVVDGLQTVDSDSIEP 6379
QY 121-----
Db 6380 IERNAPLALFSOORLWFLDRDPCGASSAYHMPMSLLLRGELDRALKAAALDRVARHES 6439
QY 121-----
Db 6440 LRTTFELHGEQPVQVITAAADSGFALAEDDLRSQPYEQASLNASRIADSEAAAPFDLROGP 6499
QY 121-----VTQ-----
Db 6500 LIRGRLLRLADDEHMLLIITQHHIISDGSVGVLINEFTALYQAFTEORPDPPLPALSIOYA 6559
QY 124 -----LGGACSP-----
Db 6560 DYAAWQRRTFTGERLAEQADLWREHLGGA--PTLLSLPTDRPRPVVQSVYRGAVPVTIDA 6617
QY 132 -----WSCLIT-----
Db 6618 ALHQRLERFCQAHNVTLFMGLLSAWSVLMTRLGNERDQVIGVPSANRGRTETENLIGFFV 6677
QY 138-----
Db 6678 NALALRVDLTONPSVAQLLEQVRQTTLAAHEHODIPEQVIEALQPPRSMHSHPLCQVAL 6737
QY 138-----
Db 6738 SLDNTSIGGELKLPGLSLHPVLOAHETAQFDLMLTLASENGALNGVIEASDLFDRSTVE 6797
QY 138-----
Db 6798 RFAQHFTLLEAMVEDVAQPVGLPLLSPAQRLASPALLOPKAVFASGLMVHORFEOFAA 6857
QY 138-----
Db 6858 AHPONTALVGRHEVSQALNRRANRLAHELLAAQGVPRPDRVAI LAERGTCQMICAVLAVL 6917
QY 138-----
```

Db	6918	KSGAAYVPLDPTYPTERTBRLGYYLLTDSAPVALLAQSACLVDLPAHSVPVLTTLGADNDSADLA	6977
Qy	138	-----	137
Db	6978	DESVDNRPEPALCLGREHLAVIYITSGSTGLPKGVLVEHGNVARELFDATAGQNFQGHED	7037
Qy	138	-----	137
Db	7038	VWTPHFSFAPDFSWBEIWGALCYGGKLVIVPSEVARSPDDFVALVCEQQVTVLNQTPSAF	7097
Qy	138	-----	137
Db	7098	RQFIQARERSPOEHALREVVFGEALDFRSLQPWARTPLSRTRLVNNYGITETVHATY	7157
Qy	138	-----	137
Db	7158	YPIQSEIDTAMPSLIGPALDCLRLDDYQOPVPGVNGEIYIGGAGVARHVLNRTEL	7217
Qy	138	-----	137
Db	7218	NABRFIADPYALQSGARLYRTGDVAHYRSDGGVWVNGRNSQIKIRGFRIELGEIEAQLL	7277
Qy	138	-----	137
Db	7278	ACPEVREANVILREDRPGKRLVAYLIAEDGAAPESALLRSQLASVLAEHMLPSAFVTLLE	7337
Qy	138	-----	137
Db	7338	TWPLTTNGKLDRAALPAPDOFATVSRDYEAPLGAJETTLAAWQELLGVERVGRQDHFFE	7397
Qy	138	-----	137
Db	7398	LGGHSFLVISLIERLRQIGLLLDVSTVFSAPTQAMAAVLAGGTAAERVAPANLIPVDCT	7457
Qy	138	-----	137
Db	7458	ALTPDMLPLVKLTQRELEHIVADVPGGVANVODIYPLSLQEGILFHHLLQSEGDAYLMR	7517
Qy	138	-----	137
Db	7518	TLATFDSRALLDKFLGALQVVINRHDIRSSLRWQGLPOPVQVVHQAQLPVIQDTPAG	7577
Qy	138	-----	137
Db	7578	EDALQMLRERTNTYHMRDLQQAPLIAAYITYTRQEKWMLALDHHLLISDNVTRLIMG	7637
Qy	138	-----	137
Db	7638	EIQAVMDGRADALPPSQPYRNFARAAVCVQAEHEAYFRQLLGDVDTTAPYGVLDVRGG	7697
Qy	138	-----	137
Db	7698	DATILRSVQDLSDDLARSIHSTARAQCVPTSVLFHAAWGLVVAATSGRDDGIGFTVLSCR	7757
Qy	138	-----	137
Db	7758	SQGTSGANHALGMFINLPMRIBLQNSVRDIVODAYOOLSGLLTHERAPLALAQRCSAV	7817
Qy	138	-----EDTG-----	141
Db	7818	DASLPMFTVILNCRHGDLVNTAGENIEDMGEEGVHFLGSETRNYPIETIAVANEANGF	7877
Qy	142	-----	141
Db	7978	SLTAQSIDGIDPHRIAYALGQVAELVNALEQDPAPLASSLEVIPOAERQLLLNDFNDTA	7937
Qy	142	-----	141
Db	7938	SDFAPAVPIHALFEDQVRNPDVALVYEDRQLSVQLNRRANHVARQLQLGVQDPERV	7997
Qy	142	-----	141

Db	7998	AI CAERSLDWIAGLLGVLSKSGAAYYPIDIPAHAPADRMKAFMLQDSOPRALLTQGALLSLPVG	8057
Qy	142	-----	141
Db	8058	TPLMLLDSAESLLAADQAFDANPVVDGLTAENLAYVIYTSGSTGQSGVMVHRSVFNF	8117
Qy	142	-----	141
Db	8118	WNVLTRTHOHCPTPATVALNAGFFDMSIKGISQLFSGHKLVIIPOLLRANGSELDDFL	8177
Qy	142	-----	141
Db	8178	EAHQVAFDSTPSQDPTLLSAGLLERSYQPVSVLLGGEAINASTWEKLRNCPTIRLYNM	8237
Qy	142	-----	141
Db	8238	YGPTECTVDATIDIRLGEKPSIGRPANVQVHVLDARGEPAPLGVAGEIHIGSGVAR	8297
Qy	142	-----	141
Db	8298	GYLNRDELSAERPFDVFPFSDAANARLYKTGDLGRWLADGTLEYMGRNDFOVKVGRFRIEL	8357
Qy	142	-----	141
Db	8358	GETENVLLAVPGIREVVVIARNDSDSDSORLVAIVCGESVAAEHLSRSELLRHLPEYMV	8417
Qy	142	-----	141
Db	8418	PSAFVQLDALPLTANGKLDORRALPAGQDALASKVYEAPQGETEATAEIKALLHLDOV	8477
Qy	142	-----	141
Db	8478	GRNDFLELGGHLLAVQLLSRLRKLGTRITRELFDAPTVRGLASLVNAAAPGEAQSI	8537
Qy	142	-----	141
Db	8538	PRANRSGRLPLSFQORLWFLDHLDDHAAAGAYVHLPMAURLTGALDTAALEATLDRLVARH	8597
Qy	142	-----	141
Db	8598	ETLRTRELVDGEPVKIAPADSRPLLKQDLRLDSDGERTSTLARLGOENATOLFDTLK	8657
Qy	142	-----	141
Db	8658	GPMLRGHLLRVADAEHVLLITLHHIVSDGWSNVLAQEVSALYAAFSOGOKDPLPALPLQ	8717
Qy	142	-----	141
Db	8718	YVDYAAWQROSLDGPALQAQIDFWRKHLEGAPSVLNLPLDRPRPAIOSYTCGMVHVFSP	8777
Qy	142	-----	141
Db	8778	ALSADLRAFSAQGSTFPFVULLAGWSMLSHLSEQTDVVVGTPPVANROHPELEPWGFFA	8837
Qy	142	-----	141
Db	8838	NTLALRVATDRETLNMLDRISLTLAAYNHODLPFEQVVSALQPTRNVSHSLPQVML	8897
Qy	142	-----FDLGVTI 148 :	
Db	8898	SLDTPPSLLQLPDLEVELLDSEHHTTQFDLSLSL 8932	
RESULT 6			
T15789			
hypothetical protein C41A3.1 - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 01-Dec-2000			
C:Accession: T15789			
R:Bentley, D.			
submitted to the EMBL Data Library, November 1995			
A:Description: The sequence of C. elegans cosmid C41A3.			
A:Reference number: Z18404			

A:Accession: T15789
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7829 <BEN>
A:Cross-references: EMBL:U41541; NID:g1109867; PID:g1109869; PIDN:AAA83181.1; CESP:C41A3
C:Genetics:
A:Gene: CESP:C41A3.1
A:Introns: 29/2; 69/3; 141/2; 192/2; 271/3; 321/2; 1000/1; 1098/1; 1127/3; 1168/2; 1282/2; 7514/3; 7596/3; 7654/3; 7716/3; 7732/3; 7769/2; 7797/3
C:Superfamily: acyl carrier protein homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:6547-6616/Domain: acyl carrier protein homology <ACP>
F:2832,5271,6580/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 30.1%; Score 233; DB 2; Length 7829;
Best Local Similarity 1.3%; Pred. No. 65;
Matches 90; Conservative 25; Mismatches 33; Indels 6699; Gaps 27;

QY	1	AAGGI-----	5
Db	151	AAVGIETACSSSLVAFHARAOIQTGETKLALVCGANHVGSRSPHSLYNHVMSPNGRLA	210
QY	6	-----	5
Db	211	AFDRSANGFVRAESFAVAVLCQKFAEENLLIHCECVGSAPNSDGKTPSLTAPNPISQY	270
QY	6	-----	5
Db	271	EVOLEAKNIDKDSVOLVTCGTGKLGDOVELTAINRSFKSDIRVMSPKSMGHGEGAA	330
QY	6	-----LHLEL-----	10
Db	331	GLIGVLSQSLYMQHGIIIPNQLHLELPSEDLGEDKSMGFNBEMELNRVAISSYFGGTNA	390
QY	11	-----	10
Db	391	CAIIEKPEKPSLVKESYAESNVFLPSAKSHESLKLQIEEYTFQFAQSDSAMEDILTYVN	450
QY	11	-----	10
Db	451	ERKTYDFAAVFGKDNIEIARKLQDGDYSLTNLQESTFEVEFGEGNEKLWLLRLMYEKN	510
QY	11	-----	10
Db	511	ETFHSTVDKYCKLAETCGFPEARTALFPFKLTLTPLTYNVSRLLISSMATFELLVQNTL	570
QY	11	-----	10
Db	571	PNKLRGKLGQIFCLAVAKVITFESAVOLIKGVVAEANLTDILGDIELKSKKIPIEIQHL	630
QY	11	-----	10
Db	631	KSTKKKILPIHISGELKETAKPNLMTFIVNGEILELDPVRKVQKLIICOLFACGDPAPVK	690
QY	11	-----	10
Db	691	FRGRIVKTPYFLKQFQWPEVQMTMTNIVDEQTNSSLSDAEISTVTVIKQFLDIEE	750
QY	11	-----	10
Db	751	DDINLLETGAVDSLTSIEMVEAFGTAVNQTMPFDLLLEAYPTILNIVDFLKLTVTPVTK	810
QY	11	-----	10
Db	811	ATTSIHKTSELSTSDINVIACDQFAGVEGKELWDLTSLRTTGKISDIRKKQCEGD	870
QY	11	-----	10
Db	871	AGLEVGLLKQDISMPDNSPFAIKOEAFLDPQHRLLNAAYNALKSGLTSIPDADFL	930
QY	11	-----	10
Db	931	AISAHSEYRALAEKHINELDERLWGTVHSMVAGRLAVLMGIRGAMIVDTTCSSVATAL	990

QY	11	-----	10
Db	991	EMAVKSIREGRKPAIVATSQLIOSSKWLKSLKTLDDHHSTNSFVDCSGFCRSDGVGVII	1050
QY	11	-----	10
Db	1051	LKTAEGDSAVIKISSAKSHHCAGVMTVPVSSISQLEEAGSFSYVEGHGTATSAGDSAE	1110
QY	11	-----	10
Db	1111	SMAYQKLGSELIMSSVKAQFGHCEVAGSLQLMKVSSIGKHGIIPISVHNILPSEHIRNN	1170
QY	11	-----	10
Db	1171	ENIRLPFAEEKIDRSAIVSFGITGKTIVVTTERVSQNLVNDNIONCYLLPVSAKTKOGL	1230
QY	11	-----	10
Db	1231	KKACLSLIEIMDNSCESLYDISTTLOKQKTNPKWRTAVVGSSHADVVLKQKFLTSEHT	1290
QY	11	-----LVAV-----	14
Db	1291	SLTNWHISTSHHSIGCSITFFHNIPEFEDHYSMFCHRLRLEFEPHSNTNTSIYHLLAVVYALIR	1350
QY	15	-----	14
Db	1351	VILKHLKLTNSFAVGFSNLIVLAAVDAAPSHYLNLDLLHAFANDDDVRMMKRIARDVTISIE	1410
QY	15	-----	14
Db	1411	NVKKLNLNGEPIITTAQAVEATIDOKVKNVRLPETTLILSPRAYEFASOLETTIDQYKL	1470
QY	15	-----	14
Db	1471	IGKEFTGGQVDFAGIFGTPIKLDLPEYPNRSFWLPIDPSVPSNEKEGKELIPKSYE	1530
QY	15	-----	14
Db	1531	FLKSKQKQHVONHVDSKIVLPGATSIRLVHOLNGKPTVELSNIDFLNKITPSEAPSVV	1590
QY	15	-----	14
Db	1591	KIBEQDGLKLVFGETAISFKLTELQNFNPIERNRLNAEVHTDNIYERFANSHTLYRN	1650
QY	15	-----	14
Db	1651	EFQVDSLKTYMGKEVRFVSMKDLIDLIDGTLQAIVCVCFPENTNDNSPFVPFTIDQLS	1710
QY	15	-----	14
Db	1711	ILNGDISOKOLHAVLYDSSGNFNGDATVYDALGNILHISNVTFKRLNGOSAPSLTSK	1770
QY	15	-----	14
Db	1771	TVDSKITKKVENEDOKRASKNNLHVWFENFGWTDIDNTTGFDFDLGLTSIQAVKLRAI	1830
QY	15	-----	14
Db	1831	KSYPNASSTCVDPYPSIDLLSGYLSTLNDPOVTETSTGEDDIOKDLTEDHKPRLAENP	1890
QY	15	-----	14
Db	1891	IGVMAAACRLPGGVSSPSSELWELLKIGKNASSRIPATRVPTNTLISGSKYGNPVEGNF	1950
QY	15	-----	14
Db	1951	ITQDVTQDPDPFFKISKSEBELIDPQORLLECEQLENSGVIETSNVGVFVGLMEKEY	2010
QY	15	-----GP-----	16
Db	2011	QDMMESSILAMLGSMMAVIAGRVNYIFGCYGPSVTIDTACSSSLVALEMAINALLDNRC	2070

QY 17 ----- 16
Db 2071 SKVIVAGVNLILNEKGGLRTNGKMLSOHGMSLSPDGRSGYGRDGCWLMLELAKPNF 2130
QY 17 -----DVFOAH----- 22
Db 2131 HYMTIOSVNVNHGGRSVSLTAPNGVAHKMLLTSVINOSPSLAIDYWEAHGTGTPGLDPI 2190
QY 23 ----- 22
Db 2191 EFNTLSSTLQNIIGSVKASLGHGEASAGTCGLLKLFLMLTYQVPTLIHFHVLNKDINA 2250
QY 23 ----- 22
Db 2251 GSIRLPIIGEDSELVSAGISSFQVSGTNAAAIAFNDNNKLEPYPIHKKYILPISAKNOI 2310
QY 23 ----- 22
Db 2311 SLDNLEKQILSVIPLTDVPICNIASALANNRSHFTIRNALIVSNSGIVNSKMEGKPHRVA 2370
QY 23 ----- 22
Db 2371 KKDRVHVKCLDSDLASLLQYDVINETYVASLKNPQSFAMKFAIKELTSLSEYIEIVA 2430
QY 23 ----- 22
Db 2431 SDGELLAVLLANGSLKWFNFKTMIELPIGSLLTAFADHDLNSTSSSIKSYOTQPESH 2490
QY 23 ----- 22
Db 2491 NLDSPMELMKLMKIYITGYDWDATVYSPVEQFIALPNYQFNKOTLWFEERLEIVDHYL 2550
QY 23 ----- 22
Db 2551 IGTIDEESEDTLILKNOISELRHPQFFKGKPLDVGTMSEIAIEALKIRNEIPFSIQNLKT 2610
QY 23 -----QED----- 25
Db 2611 ELITLTPAWLETVNTRNEDEGFGVNSAYIDQGRUFLSNASSVEIQIEVPAVEQIPDK 2670
QY 26 ----- 25
Db 2671 VVYLKECPNAVIRRHNMVYVDSRAEQSPFRRTANIVLNEIGFATPSPDMFIELGVLP 2730
QY 26 ----- 25
Db 2731 VHYMYOVDDGALWQFMISQDKRVLSNIYVLKDAKGLEIPTIRMHKKSTLLSSQEAIVA 2790
QY 26 ----- 25
Db 2791 AKTLQMAVRHKVCLAVGDVIESGLDIDESQLSTGFSSELGIDSLATVDLLNRLNKYFPEI 2850
QY 26 ----- 25
Db 2851 ELTTSOLFNDPSIIDLSIMEIOLNEKGITEPSEBPTKTSRGRKLSIPAVRAQVLAQI 2910
QY 26 ----- 25
Db 2911 EFVENYNSKQKEVQAEAPSSSECSNHLSESDATVDTRTEIRRKVSLAVFDLATETLSAED 2970
QY 26 ----- 25
Db 2971 LOSKGFTELGMDLSIVDFVNRNDKYFPDDEITASDIPDYPTVDELSDHIVRKKSSVP 3030
QY 26 ----- 25
Db 3031 PAASEIMKETMNGISTSVDAETHKLENLSQSFMLENQNSINPTLXMIWSNOTIKLVKPS 3090
QY 26 ----- 25
Db 3091 DGNFLFELNANGQEQEIKQKFTGPNNIIDLKGFHEGSTETLYMNSLLNLVKSISKLEIQ 3150
QY 26 ----- 25

Db 3151 CRGVSQEFGLGNSISRAFMKTVAEKNPLISFANYQNVQVSFVSDSPITGNWLTGG 3210
QY 26 ----- 25
Db 3211 LSGIGLEIGKFIANGAENVILISRROPTAKALRDILSTELTHIGLARKTIVLKIKNIS 3270
QY 26 -----TERYV----- 30
Db 3271 AKLIVFOSFKLSFSTKYYISDLFSKKVTPFPYNLLQSKCFKHFHWFLEWLEVYLIKQWTV 3330
QY 31 -----LTNLNIG-----AELLRD----- 43
Db 3331 HTAADINDKEKLIARELTCLNVGITGIHSAGVLKDSKIERONKESPNQVTPKANGPHV 3390
QY 44 ----- 43
Db 3391 LEETEKHFNYKIENFIMMSSFTAACGNEGQLVGSVNAYLEYOVQRRRRROCKSGCAIOMG 3450
QY 44 ----- 43
Db 3451 NWIDTGMATDENVRKFLANLGLQHNKDALKYLKACILTKPELIMVANIDWVILKNRK 3510
QY 44 ----- 43
Db 3511 DLPKDLINTGILPDEFTGKMESEFPPLSNGDPFEKVSMMFSEDEEVELELIKEKVSSIL 3570
QY 44 ----- 43
Db 3571 MCSPTKLKNNKNIMDGLSKLIVEFLNFNSTFKISVNLSDAYNHPTLEKLAHAHIFOM 3630
QY 44 ----- 43
Db 3631 TIVDHPVNSVKSEBIFKSTDFCPIFGINIPFDNKNDFDPAKSTAVKLENGEQLPTAGKY 3690
QY 44 -----PSLGAQ----- 49
Db 3691 AVSVVGSIRDVYSKIEAAPQIKLCQESSKCVLMLTGGSGOYPMHGRQLVENYEIFRT 3750
QY 50 ----- 49
Db 3751 TLQSLCKCDEYLQGVSLWEILFNTDHYKLLQLTQHMOPIMFCGYATAQLWLSLGI 3810
QY 50 -----F 50
Db 3811 DYLLGHVSELVAGVLAGIMSIEDGLRLIVERGKAMENIAGLGALLAVQREIADVLKRP 3870
QY 51 RV-----HLVKM-- 57
Db 3871 KVSATINSKQVVFAGTKSVLDAALAFVKQGGKQATVYVQYPPHSHNLQETHLVLSURQ 3930
QY 58 ----- 57
Db 3931 CLADIKFSAGRTPLVSNVTGOINTFSEAYIVKHTSAVKFVDCVETLQAKGVTVMIDAG 3990
QY 58 ----- 57
Db 3991 SAAVLATFVKRIIOPTELSKHRIVOTCKEKESVDNLVQACLEEQSLPISWTTLYCGG 4050
QY 58 ----- 57
Db 4051 RNADERLVEFPMTNDIINKDEFELLEQHQLNGKIVVAGAYQLFKIDQLVKLKAAGMELM 4110
QY 58 ----- 57
Db 4111 LKNVFLKPMYIEDNREYQIQWNSDMTIELIVNSVIVCSLEVEPQNSVLKLETISENKP 4170
QY 58 ----- 57
Db 4171 FEVHDFVETLFRNGLOVDSGFRRIESARRSDKRCFSQIKSPPFAWPLIDSAMHSITASV 4230
QY 58 ----- 57

```
Db 4231 PRRDCYFLPVAMGSVTMKTDSFTPLNLHAQTVITSETDKFIQNVALLAGDTPICEVR 4290
Qy 58 -----VILTEP----- 63
Db 4291 NMTIVLUKTEPVHTRIPNSIETVETPPKSEIEIVGFDISLPYNOISENENWHLKNTV 4350
Qy 64 ----- 63
Db 4351 KQLHNRSLKQDHARVALLDSADARYWDPYEGIRPSEAKFIDPOORLLCSVAKLLDSSL 4410
Qy 64 ----- 63
Db 4411 ITSLSNTGVFIGSANEFSHIVAYGYKOPRAEWSGGTNSALAGRIAHWLKLGPPVT 4470
Qy 64 ----- 63
Db 4471 LDTACSSSFYALSACDALRTGQCEYAIIVGTNVLVMEHMTDVLQNAKMTVDDFCKAFDV 4530
Qy 64 -----EGAPNITA----- 71
Db 4531 DANGYKSEAVCSMLLTSPNIDSVATITNYATGHNGTSSSLFTPNGLSQLEVWQATNP 4590
Qy 72 -----NLTSSLLS 79
Db 4591 LEKILEIQTCTGTGLGDPIDIEINAISKLVSACKIGSVKSNIGHTGGSLVSLCSLSMS 4650
Qy 80 ----- 79
Db 4651 FRSKYRVAQLHLKPTNSIKTNMKICRFIGEDADENNILINFGFTGSNCNVVLKPKNA 4710
Qy 80 ----- 79
Db 4711 ISEHFVSSEVPYILLSSHSAKSLOKYVOVLCEFISNSAKSLDHIMMSLFQKKIHVRQF 4770
Qy 80 ----- 79
Db 4771 IIFNFKRIAVTSLDGEVFRDERLEKLKHPCSVFLKEGVVHFDKDSFORVDLPSIVFN 4830
Qy 80 -----VC----- 81
Db 4831 NTLHWALDSYRDEIDRHESQMSPKNIFYEKVLLETMPQONPKSVKVCIGRDLIPKIE 4890
Qy 82 ----- 81
Db 4891 IDEVSSFANGIIIVPHIPISNIFEYLKAKVWSLISRNQNVFIICCENGTSHTWTGTL 4950
Qy 82 ----- 81
Db 4951 RSLASERKMPYKVSIDKVDALKLEFNHEYMFEAIFYSKRGYVERLRKVNPTKFAPOY 5010
Qy 82 ----- 81
Db 5011 EKCLISGTTGIGSAIINELPKSSVITRKNIASDEGKTFELSSDITRLDISHKFNVVFH 5070
Qy 82 ----- 81
Db 5071 LAGIVNNSLHENVKRDSDLENVSIKLOKAKNLMKCCDSETHFVSSSIANVLGSGOSNY 5130
Qy 82 -----GW----- 83
Db 5131 AFSNGLVTSFLETSTKSTIIHWGPWKDVGMQAQPERREIVKQIESNGWKLLPNQDAISV 5190
Qy 84 ----- 83
Db 5191 PYOTFMETHEOIIVFDGGDFTIVARQPHLOKLLSEVVEKTKVKEEIKKKSUNFEEIFEI 5250
Qy 84 ----- 83
Db 5251 VGITDISSKLNIPEMDLIGDISLCHMENLRYSLNKNFDLELTVSEMFENATYQKLOTYYVETL 5310
Qy 84 -----SQTINPEDDT----- 93
Db 5311 RKAKHSESLSHDRVSSQVSNKEDDTRVAVIGWSAEFGSSNIHEYWENLMDGICSTGNK 5370
```

```
Qy 94 -----DP----- 95
Db 5371 YLLKNPFGDNKFPNLTDEDARVLDPOVRKFIOHAYLALENSGYVQKHRELRCGVFAGAE 5430
Qy 96 ----- 95
Db 5431 PSDYGRADDDHDAWKRLFVMNMNSYLASYASYCLDLKGEAVSVYSACSTALVAVANAVKS 5490
Qy 96 ----- 95
Db 5491 IQSGMDYALVGAASIAEVSGALSGDFDQKKTWFSKSGVCRPFDDKSEGIVRSGVGCFFV 5550
Qy 96 ----- 95
Db 5551 LKRYSOALLDNDNVHVIKDFAINNDGHSRASFMAPNAGQLKCMTDVLFARFTNKERKI 5610
Qy 96 -----GHA----- 98
Db 5611 SFVECHATGTTLGDITIEMNSLRtaySFKNKLAIGCKCANIGHAYASGLAALVKCAKMLQ 5670
Qy 99 ----- 98
Db 5671 TGIIPQVNFSEPRDGMGOFTVNGKKSTISONSLISIDSFGIGGTNVHVIFFPARSQE 5730
Qy 99 -----DLVLY----- 103
Db 5731 VWKISSENILYDMIPISAKTEYSLOHTSEAIKSYLOTDNTNKIAQCSSTFIHSRVPMSDR 5790
Qy 104 ----- 103
Db 5791 TYLSVNNNNELLKIRTNKKTWFKNGSPKIALPPAQIOFTNLPNEYLNKSAYRREVEY 5850
Qy 104 ----- 103
Db 5851 LCELASSFGIPSLLEGILYPTKNFDHLIHATOPAQIAIFVOCMAIFKAIKNVFNPTCLIGH 5910
Qy 104 ----- 103
Db 5911 SVGEYAAAVISGALKTEEALKLLIKRSELJGKTEKARMLVMWNYEKOLPSHVHVSAIIDA 5970
Qy 104 ----- 103
Db 5971 NTKCVVPVETIONLEKYFINNHIKYRNIEHTKHGPHSKMPCISKEFEFFCESFATKVP 6030
Qy 104 ----- 103
Db 6031 IPMISSITGSEIKIPDSKYCTMHLTNPNVLELVVDHIMKLDIDIIVEVGTGVLNLLAK 6090
Qy 104 ----- 103
Db 6091 RNSKIVVVPCTGTHKPKISLGECIGQLMSNGVDIRKLTPLKAVDGOVPGYCFCDERQFGE 6150
Qy 104 ----- 103
Db 6151 KNQKISQKLNRPYSYKESWKPVSSLEETKPVRTVCDGSLNGISEPVLVLLRKSRDIH 6210
Qy 104 -----ITRDL----- 109
Db 6211 KNYFDVELFLKSPRCSTVVFGMDNSPSVHLSGLIRCYQLVSRIDLKYVENPENIAMS 6270
Qy 110 -----ELP----- 112
Db 6271 IVIOKVLTLANGLYFRIDSTCLCKHGFRTEIPSRLLHPVSRGHALVFGANGFIGSIVFRLL 6330
Qy 113 ----- 112
Db 6331 QEMGMNVIPISRASIPSCDITNIKDVONVFKSLGFKKFSVINCVGVETSAKMNKTSLEQ 6390
Qy 113 ----- 112
Db 6391 EIVLSPKTFGSVNILKCLEEFSIEVDKLVNFSSLSVVPLLGNFDYASANCFVEALTQOG 6450
```



```

Qy 113 -----DGNROVRGT----- 122
Db 6451 SKYIKOFLTLSPLEGRMYESSSTKHNLSQILMFOEILTNNLKTAFESNMNGIVFV 6510
Qy 123 ----- 122
Db 6511 SSVNPADIATKSIAPHRDENGSEVKAVESLKLPKSTCEPFAIEIWKETLGISILNDAN 6570
Qy 123 -----QLGG----- 126
Db 6571 PNFFSLGGDSLSALQVVKVQKTDRIVDVNDLFDNPTLOEFTKFMNLTTEKFAGNTND 6630
Qy 127 ----- 126
Db 6631 KISYDAIPLTNSOTOMFLROIQTTSKYNLIPKITISYETKFWBFLKYSLSLHSLIAYOPS 6690
Qy 127 -----ACSPT----- 131
Db 6691 YRTVFKSGNSPYOYICSLTESFDFDKRCNNLNAISHEPNHLFEICKSTPLRVRVAECD 6750
Qy 132 -----WSCLITEDT----- 140
Db 6751 NSRIHIVNQHHILTDGWSMTVLSDTVSSLYAAYRGTSFSPSKTKQTISQVAMGKSSGD 6810
Qy 141 ----- 140
Db 6811 IKALEYQNTYHTIIPYDSEGTGNTSPSYVRISKLIPIKWKLVGLSKLYNTTMYNLAL 6870
Qy 141 ----- 140
Db 6871 SVPCDAVRSFTGQADILLAYAGRNADNSSELIGYFMNNAIPKTSIPFELRLLEILNIV 6930
Qy 141 -----G 141
Db 6931 LNSLEKRSFATIPFYQMVQNRKLNELISLFFNFRQKLDYPTVSMFGAKCEIEHLNNA 6990
Qy 142 FDLGVTI 148
Db 6991 FDFSFTI 6997

```

RESULT 7
T21371
hypothetical protein F25H8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21371; T24896
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19413
A:Accession: T21371
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <W1>
A:Cross-references: EMBL:269360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: Clone F25H8
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19949
A:Accession: T24896
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <W12>
A:Cross-references: EMBL:269361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone T13H10
C:Genetics:
A:Gene: CESP:F25H8.3
A:Map position: 4
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 81

Query Match 29.6%; Score 229; DB 2; Length 2165;
Best Local Similarity 3.7%; Pred. No. 1.2;
Matches 61; Conservative 20; Mismatches 65; Indels 1522; Gaps 8;

```

Qy 3 GGILHL----- 8
Db 5 GGSFHLQPVVAAILLVWCLVYALQSGGTISEFSSDVLFSRAKYSGVPVHHSRWRDA 64
Qy 9 ----- 8
Db 65 GIHVDSHHIVRRDSYGRGKRDVTSTDRRRRLQGVARDCHACHLRLSRDDAVYIVHLH 124
Qy 9 ----- 8
Db 125 RWNQIPDSHNKSVPHFSNSNFAPMVLYLDSBEEVGRGMSRTDPDCIYRAHVKGVQHSHV 184
Qy 9 ----- 8
Db 185 NLCDSEDLGYMLALPSGIHTVEPIISNGTEHDPGASRRHQLVRKPDPMHFKSFDHLNS 244
Qy 9 -----ELLVAVGPDPVFOAHQEDTERYV 30
Db 245 TSVNETETTVATWQDQWEDVIERKASRRANSWDHYVEVLVADTKMYEYHGRSLEDYV 304
Qy 31 LTNNIGAEELRDPSLGAQFRVHLVKMVLITEPEGAPNITANLTSSLLSVCGWQSQTINPE 90
Db 305 LTFSTVASIYRHQSIRASINVVVVVVLKLTENAGPRIITQNAQOQTLQDFCRMQOYYNDP 364
Qy 91 DDTDPHADLVLYITRFDLELPDGNROVRGTOLG----- 125
Db 365 DSSVQHHDAVAILTRKDCRSQKCDTLGLAELGTWCDMOKSCAIIEDNGLSAAFTIAH 424
Qy 126 ----- 125
Db 425 ELGHVFSIPHDDERKCYTMPVNVKCKFQSTKPDKTOFQNNFHIMAPTLEINTHPWSWP 484
Qy 126 ----- 125
Db 485 CSAGMLERFLENNRGOTQCLFDQPVERRYEDVFRDEPGKKYDAHQCKFVFGPASELC 544
Qy 126 ----- 125
Db 545 PYMPTCRLWCATFYSGMGCRTOHMPWADGTPCDESRSMFCHHGACVRLAPESLTIDG 604
Qy 126 ----- 125
Db 605 QMGWRSWGCSTCGGVQKGLRDCDSPKPRNGKCYVGORERYRSCNTQECPDWTPQY 664
Qy 126 ----- 125
Db 665 REVOCSEFNKDIGIQGVASTNTHWVPKYANVAPNERCKLYCRLSGSAFYLLRDKVVDG 724
Qy 126 -----GACSP----- 130
Db 725 TPCDRNGDDICVAGACMPAGCDHQLHSLRRDKCGVCGGDDSSCKVVKGTNEOGTFGN 784
Qy 131 ----- 130
Db 785 EVMKIPAGSANIDIRQKGYNNMKEDDNYLSIRAANGFELLNGHFQVSLARQOIAFQDTVL 844
Qy 131 ----- 130
Db 845 EYSGDAIIERINGTPIRSDIYVHVLSVSGSHPPDISYEYMTAAVNAVIRPISSALYLW 904
Qy 131 -----TWS-----CL-----IT 137
Db 905 RVTDTWTECDRACRQGSQKLMCLDMSHROSHDRNCQNVLPKQATRMCNIDCSTWIT 964
Qy 138 ED----- 139
Db 965 EDVSSCAKCGSGOKRORVSCVMEGDRQTPASEHLCDRNSKPSDIASCYIDCSGRKWN 1024
Qy 140 ----- 139
Db 1025 GEMTSCSTCGSNGMKHRKSYCVDDSNRRVDESLLCGREOKATERECNRIPCPRVWYGH 1084

```

QY 140 ----- 139
Db 1085 SECSCDGGVGRHAQCLDAADRETHTSRCGPAQTEHCNEHACTWQFGWSDCSAKC 1144
QY 140 ----- 139
Db 1145 GDGVQYRDANCTDRHRSVLPPEHRLKMEKIITKPCHEKSCPKYKLGEMSCQSVSCEDGWS 1204
QY 140 ----- 139
Db 1205 SRRVSCVSGNGTEVDMSLCGTASDRPASHQTCNLTGTCFFHWNTDWSACSVCSCGIGHRETR 1264
QY 140 ----- 139
Db 1265 TECIYREQSDASFCDGDKMPETSQTCCHLLPCTSKWPSHWSPCVTCGSGIOTRSVSCTR 1324
QY 140 ----- 141
Db 1325 GSECTIVDEYFCDNRTRPRLKKTCEKOTCDGPRVLQKLOADVPPIRWATGPMWTACATCG 1384
QY 142 ----- 141
Db 1385 NGTORLLKCRDHVRDLPDEYCNHLKDEVSTRNCRRLDCSYWKAWEBCPATCGTHVQQ 1444
QY 142 ----- 141
Db 1445 SRNVCVSAEDGGRTILKDVDCDVQKRPSTSARNCRLEPCPKGEEHIGSWIGDKSCAS 1504
QY 142 ----- 143
Db 1505 CGGWRRRSVSCTSSCDETRKPMFKCNBELCPPLTNNSWQISPMWCHSVSCGGVQR 1564
QY 144 ----- 143
Db 1565 RKIWCEDVLSGRKODDIECSEIKPREORDCEMPPCRSHYHNKTSASMTSLSSNSNTTS 1624
QY 144 ----- 148
Db 1625 SASASSLPILPPVSVWTSWSACSACGCRGTRKRVVCEVNPVSLNVTV 1672

RESULT 8
T36248
CDA peptide synthetase I - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 01-Dec-2000
C:Accession: T36248
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21602
A:Accession: T36248
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7463 <SAU>
A:Cross-references: EMBL:AL035640; PIDN:CAB38518.1; GSPDB:GN00070; SCOEDB:SC563.03c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: cdaPST; SCOEDB:SC563.03c
C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:516-1074/Domain: acetate-CoA ligase homology #status atypical <ACLI>
F:1090-1158/Domain: acyl carrier protein homology <ACP1>
F:1715-2184/Domain: acetate-CoA ligase homology <ACL2>
F:2200-2268/Domain: acyl carrier protein homology <ACP2>
F:2804-3249/Domain: acetate-CoA ligase homology <ACL3>
F:3265-3332/Domain: acyl carrier protein homology <ACP3>
F:4323-4746/Domain: acetate-CoA ligase homology <ACL4>
F:4762-4830/Domain: acyl carrier protein homology <ACP4>
F:5363-5786/Domain: acetate-CoA ligase homology <ACL5>
F:5802-5870/Domain: acyl carrier protein homology <ACP5>
F:6401-6868/Domain: acetate-CoA ligase homology <ACP6>
F:6884-6951/Domain: acyl carrier protein homology <ACP6>
F:1122,2232,3297,4794,5834,6916/Binding site: phosphopantetheine (Ser) (covalent) #statu

Query Match 29.6%; Score 229; DB 2; Length 7463;
Best Local Similarity 1.5%; Pred. No. 89;
Matches 86; Conservative 22; Mismatches 39; Indels 5712; Gaps 25;
QY 1 AAGG-----IL---HLEL 10
|||
Db 1576 AAGQAGLDLAVFSTELFDAGTVRALTERLLLLLTQVAADPGVRLGDDLVLLPGEHDL 1635
QY 11 LVA----- 13
|||
Db 1636 LVAANRGTAPASPAPSPASPAPGTATILPELPEROAAHPDRTALTTFCTSL 1695
QY 14 -----VGPDV----- 18
|||
Db 1696 SYAELNARNLARLLTARGIGPDALVALALPRSAELWALLAVKSGAAYVPLDPGYP 1755
QY 19 ----- 18
Db 1756 DRLAHALSDSPAALLTDRATGRLPAHEVPRIVLDAPADGTTGGDPADAHPTDLA 1815
QY 19 ----- 18
Db 1816 QGSRVRPLDPRDTAYVIYTSGSTGRPKGVAVPHGNVVRLEFSATAPWFGPDEHDWTLFHS 1875
QY 19 ----- 18
Db 1876 YAFDFSWMELGPLLHGRLVVVPHVDTRDPAAPALLARERVTVLNTQTPSAFHQLAAAD 1935
QY 19 -----FOAHQEDT----- 26
|||
Db 1936 RENPTELALRTVFGGEALDLSRLADWYERHAEDPALVNMYGITETTTHVSHFALDRAT 1995
QY 27 -----ERYV----- 30
|||
Db 1996 AAASSASTGVNIPDLRVVYVLDRLRPTAPGVTEGMYVAGAGLARGYLGRPALTAADRPFA 2055
QY 31 ----- 30
Db 2056 DPYAALFGERGTRMYRTGDLARRRTDGLDYLGRADQVQKIRGPIEPGEIAVLAHPA 2115
QY 31 ----- 30
Db 2116 VDDVAVVAREDVQDPRLVAVVYVTGATARALHDHAAGLHPDHMLPSAEVTLTDLPLTP 2175
QY 31 ----- 30
Db 2176 NGKLDTKALPAPAHAGOVTRAPRGPREEILCALFAEVLGVPRLTVDSDFFDLGGHSLLA 2235
QY 31 ----- 30
Db 2236 TRLAGRIRGTLGVLSVRRLFETPTVAGLSAALDGAERSGTGTAGERPERPLPSYAQR 2295
QY 31 ----- 30
Db 2296 LWFHOLEGSPTYNITGALRLTGDLPGALRAAFQDVVTRHESLRTVFSDEGHGARTV 2355
QY 31 -----LTLNIGAELLR----- 42
|||
Db 2356 LDAAGVRFELPVADVSEDRDLARLEEAARHCFLTDDIPVRAELFRLGAREHVLLMVHH 2415
QY 43 ----- 42
Db 2416 IAGDWSLGLIRDLAAAYTABARRAPDWPLPVQYADFALWQRAALGDAADATSPAGR 2475
QY 43 ----- 42
Db 2476 QLAHWKEALAGLPDRLELPADRLPAVASHRGGRVPLTPAPLHSGVAELARESRTSVFM 2535
QY 43 ----- 42
Db 2536 VLOQALAAALLTRMGAGEDVPLGTVPVAGRDDADVQVGFVNTLVLTDTGGNPTFRALL 2595

QY 43 ----- 42
Db 2596 DRVRTDLTAYDHQDLPEHLVDVLSRSLSHNPLFQVLLSLDTPQODALAALSATGLG 2655
QY 43 ----- 42
Db 2656 VRLNVTTGVAKLDLALIEAHRDADGAPAGLVGAAYSAIDLFDGTVTLIVERFLRLD 2715
QY 43 ----- 42
Db 2716 ALVADPSRRIGDGDVGLPRERERVLTENDTPRRPVQGTFFADHVARHAAERPGHLAVETA 2775
QY 43 ----- 42
Db 2776 GAAAPGALTYGELNERANRLAALLARGAGPERFVAVALPRSADLVLSALAAFKAGAAY 2835
QY 43 ----- 42
Db 2836 LPVDPANPAERITHLVSDMAPTLI VTTSSALASLPDGTGTPVLLDTPETAATLAALPGHD 2895
QY 43 ----- 42
Db 2896 VTDADRVPPLRPEHPAYMIYTSGTTGRPKGVVVTHTGLPGLDIFTRDCAAGPGSRILQH 2955
QY 43 -DPSLGAQF----- 50
Db 2956 LSPSFDASFELAMGLLTGTLVAPPETTPGPELAELATRAHAATHLSLTTSVLGLPPD 3015
QY 51 ----- 50
Db 3016 SLPDGLTLVGAETPPELVERWSPORTMLNSYGTPTTVCSTMSGPLSGPAVPPIGSPV 3075
QY 51 ----- 50
Db 3076 ANSAVVYLDALRPVPPGVELYAAAGAHLAGYHRRALTAERFVANPFGEPSRLYRT 3135
QY 51 -----RVHLVKWVL----- 60
Db 3136 GDLVRMRPDQLEYLGRADTVQKIRGLRIEPTIEIAVITERPHLARAAYVIREDRPDRR 3195
QY 61 ----- 60
Db 3196 LVAYVVPFGATVDTAELRAALRETLPDHMIPTAFVVLDALPLTLNGKLDKRALPADYS 3255
QY 61 -----TEPEG----- 65
Db 3256 ARTSGRAARDPRERLLTALPGEILGVEPAGVDDGDFDGLGDSILSLQVARARAAGLVLS 3315
QY 66 ----- 65
Db 3316 VRDVFHQTALLARSAAAAPAGDRTARDSDVPADGPAPRTPMGMWFAALGSDLAAFNOS 3375
QY 66 ----- 65
Db 3376 LVLRVPAALDPDLDLTALRAVLDRHDALMRVADDWTIEIPPPGSVTPADCLVRFDVGL 3435
QY 66 ----- 65
Db 3436 DEAAVRSVTEQARTARDRLAPADGRMLQAVMLDRGADRGDGLLLLVANHLVVDGVTWRL 3495
QY 66 ----- 65
Db 3496 VPDLAAAYAGETLAPVGTPIRHWALSLSDLAGOPRTEELDHWHSVLGDTPTHRLDPAR 3555
QY 66 ----APNITANL-----TSSLLS-----VC----- 81
Db 3556 DTHATAGEITAEADADTTEALLTWPGVCHATVNDVFLSTFALAVAGWRGRGEDADAPV 3615
QY 82 ----- 81
Db 3616 VLDLESHGRHEEAVPGVELSRGTAGFTSMYPVRLAPPAGAGSGSALRALKAVKEQRTV 3675
QY 82 ----- 81

Db 3676 PGDGLGVGLLRLNPRTRAALALPLPEFGFNYLGRIGQGTDEAPWTIEGGDVAGIDGA 3735
QY 82 ----- 81
Db 3736 MPLAHPVDVNAVARETADGTRLRARWTYSRTALEPEDTORLADTWFRLLRRLVEEAROPG 3795
QY 82 ----- 81
Db 3796 AGGLTPSDIAHPALAQDEIEDLEHTVPGLODILPLAPLQSGFLFLNLYDENARDVYVQL 3855
QY 82 ----- 81
Db 3856 AFDLEGSFDCGTRMAAGALLRRHANLRAGFRQTATGTWVQVPAELEPDWRCDLTDRA 3915
QY 82 -----CWSQ----- 85
Db 3916 DEERDAEAGRLAAGDRERRFDLTPPLMFTAIRLSADRVLRVMTNHHILLDQSWPLLL 3975
QY 86 ----- 85
Db 3976 WOELTELYVSGGDPVSLPPVPRPYRDHLAWLGARDRDAARDANRRSLSGLDEATLLAPDAG 4035
QY 86 ----- 85
Db 4036 PAEAAPLGIPFGLDRDATAALSAWARGVTMNTVQCAWALAAQATGRDDVVFGATVS 4095
QY 86 ----- 85
Db 4096 GRPPELPGVESMIGLFINLTPVRARLDOABPLGDLFRRLONEQARLLDHOHPLGLADIQHW 4155
QY 86 ----- 85
Db 4156 AGHGLFDTAMVFQNYPVSAADTTSRQLDGLRVAGYDAVESTDPVNLVAHTRDDALRLRL 4215
QY 86 ----- 85
Db 4216 DYRADACAGDLVRLSLADRMLRLEALVTDSRPAVHLDTLDPVRRVRLVEMGAPTOLP 4275
QY 86 ----- 85
Db 4276 GTPLHELISEOARLTPDAVAVVCDGTSITYAELGGGANLARHLGEGLAGEDFVAIALA 4335
QY 86 ----- 85
Db 4336 KSLDAVISMLAVLKTGAAYLPIDPDYPAERITYMLDDARPALTLTEPVVVERYTGHSVTA 4395
QY 86 ----- 85
Db 4396 VTDEERRSPWSARHAAVMIYTSGSTGRPKGVVIEHHALATYLHRAARTYLAAMTGVTVLHS 4455
QY 86 ----- 85
Db 4456 PLAFDLTITALWTPLTSGTVHLTSLESDTOPSLIKATPSHLPLLTLLPETASPSHTLI 4515
QY 86 ----- 85
Db 4516 LGGEALHTDHLVTWTHQHVQIINAYGTESTVNI TDHVGEDTPDGPVPIGRPFANTQ 4575
QY 86 -----TINPE----- 90
Db 4576 VYVLDALRPVAPGVGTGELYLAGEQLARGLPALTAERFTANPHSSTPGARMYRTGDL 4635
QY 91 ----- 90
Db 4636 AHWNHHGLTYDGRADHQLKLRGHRIEPEIEATLTAQTGITQATVOLREDQGDORLVA 4695
QY 91 ----- 90
Db 4696 YLVVNDSTEYDEKTVRDALTSALPDYMWPSALVTLDALPLTPNGKLDRTLAPAPAYSAST 4755
QY 91 ----- 90

Db 4756 AGRAPRTPREVLCTLFAEVLGVDLVTIDNFFDLGGHSLLATRLVSRTRTALGVLSIR 4815
QY 91 ----- 90
Db 4816 QLFETPTVAGLAELDASGTVRTALTARPRERIPLSYAQORLWFLHQLEGPSATYNTVL 4875
QY 91 ----- 90
Db 4876 TLRGGALDVALRAAISDVVARHESLRTVFTEDERGAYQIVLPVEAASFTPTVDVAEE 4935
QY 91 ----- 90
Db 4936 EIGRLDEAVGHCFDLAQELPARTSLFRVSEHREHVLLLLIHIIASDAWSRAPLAQDLTAA 4995
QY 91 ----- 94
Db 4996 YAAVRSEAPMWAPLTVQYADYALWQOEILGDDTDADSLAGRLAYWKQQLAGLPEQDL 5055
QY 95 ----- 94
Db 5056 PTDPRPAVAGYSGDRVPFTVPTLHTRLTTELARATNTSAFMVIQAAVAVLLTRLGAGED 5115
QY 95 ----- 94
Db 5116 IPICTPVAGRTDAAADDLIGLINTLVLTDTSGDPTFRLLDRVRDTLAAVAHQDLPF 5175
QY 95 ----- 94
Db 5176 ERLVEALNPARTLSHHPLFQVLLTFNNTDHEGALKDISELPGLTVALREVORTSSKFDLS 5235
QY 95 ----- 94
Db 5236 FGFAESFDSRRPOGIEBAALDFSTELLDRRSAQAIADRFLRLVLEAVTTAPDRPIGAVELM 5295
QY 95 ----- 94
Db 5296 DPAERERVLVWNGAPTQLPCTPLHELISEQARLTPODAVVVCDGTTLYAELDRRANQL 5355
QY 95 ----- 94
Db 5356 ARHLLGEGLAEDPVAIALAKSLDAVISMLAVLKTGAAYLPIDPDYPAERITYMLDDAQ 5415
QY 95 ----- 94
Db 5416 ALTLTAPIPASYSRPTSEITDVERSSPWSARHAAWITYTSGTGRPKGVVIEHHALAT 5475
QY 95 ----- 94
Db 5476 YLHRARNTYTAMTGVTLHSPAFDLTITALWTLPTAGTVHLTSLBEEAEVQPSLIKATP 5535
QY 95 ----- 94
Db 5536 SHLPLLTLPETASPSHTLILGGEALHTDLATWRTQHPGAQIINAYGPTSTVNTIDHH 5595
QY 95 ----- 94
Db 5596 VSEDTPGVPPIGRPFANTQVVLDSALRPVAPCVTGELYLAGELARGYLGRPALTAER 5655
QY 95 ----- 94
Db 5656 FTANPHSSTPGARMYRTGDLAHNNHGHLYTDGRADHQIKLRGHRIEPEGETTLTQAQT 5715
QY 95 ----- 94
Db 5716 ITQTTVOLREDTPGQORLVAVLVNDSTEXDEPTLRDALASALPDYMRPSAYVTLDALPL 5775
QY 95 ----- 94
Db 5776 TPNGKLDRTALPAPAYSASTTGRTPRPREIILCTLFAEVLGVDLVTIDNFFDLGGHSL 5835
QY 95 ----- 94
Db 5836 LATRLVSRARTALGVLSVRQFFETPTIAGLSGAFDRAGRARAALTARPRERIPLSYAQ 5895

QY 95 ----- 94
Db 5896 QRLWFLHQLEGPSATYNIPTTLRLTGTLDLTALQSALNDLLARHESLRTYTTEDGEGPRQ 5955
QY 95 ----- 94
Db 5956 VIHAMEPGMLPLGVVDTGEGLDAMLSAGVHHAFDLTAGIPVRATLFRISEQHVLLLLII 6015
QY 95 ----- 94
Db 6016 HHIATDANSRTPLCHDLAAAYSARCAGDYPAMEPLPVQYADYALWQREVLGDEGDADAPA 6075
QY 95 ----- 94
Db 6076 GRQLAYWTRQLADLPEQLDLPDRPRPAVASQDGRVAFSLDADLYVRLTELARATHSS 6135
QY 95 ----- 94
Db 6136 FMVQAAALAVLLTRLGAGEDIPIGTVPAGRTDDATENLVGFFVNTLVLRNDTSGNPTFRE 6195
QY 95 ----- 94
Db 6196 LLEETRTDLAAVAHQDLPFERLVEALNPARTLAHHPLFQVMLLILSTAETDPDASLALPG 6255
QY 95 ----- 94
Db 6256 LRVGAERSRLGAAKVDLAFALAEVRDGEGRSTGLTGALDRTDLFORSTARSILVERFVRT 6315
QY 95 ----- 94
Db 6316 LEAVVADPGVRLSRVPLTGSERRSLDRGTGPLEGLDATLPELFAEQALRTPGAPALV 6375
QY 95 ----- 94
Db 6376 RGGTTVSYAELDLRNLRLARLLRQQQVRPGTVMMLMERSPAHVATLAIKAGAYVPL 6435
QY 95 ----- 94
Db 6436 HDITYPLDRMRHVADTAATLILTDRAEAAARAGQLGARVMVWDEFGAAPSGSEADAAPGTG 6495
QY 95 ----- 94
Db 6496 TGTGTSGRSGYDDAPEVGLRPQDLAYVMYTSGSTGVPKGVAVTHRGVVDLVRDHCHWRPG 6555
QY 95 ----- 99
Db 6556 VHERVLLHAPHAPDVSCYEMWVPLVSGGTVVVAPPGLHLPAAITDLITAHDLITAGTAG 6615
QY 100 ----- 102
Db 6616 FFRVAAEAEPCFAGVREVLTGDDVVS PAAVARVLAHHPRIVLRHLVGTETTLCVTOHE 6675
QY 103 ----- 102
Db 6676 VTAPYEARGSLPVGRATGNTRAYVLDRLYLPQVPAGVPGBELFISGGLARGYLDRLPDLTCE 6735
QY 103 ----- 106
Db 6736 RFVADPYGSGGERMYRTGDLVRNNAAGELEYLARADDQVKIRGFRVELGELIEAVLATRPE 6795
QY 107 ----- 106
Db 6796 LAQAAVVVREDRPGDRRLVGVYVVAAGRDEVPD DALRAFSRQALPDYMWPSAFVVLGTL 6855
QY 107 ----- 109
Db 6856 PLTANGKLDRLKALPAPDYGAASTGRAARTPABELLCTLFAQVLGLSAGVDDGFFDLGGD 6915
QY 110 ----- 109
Db 6916 SILSIQVSRARAAGLALAVRDVFEHQSTARLAAALTDRODAASVPEAEAVPPYGPAPLT 6975

Qy 110 -----ELPDGNQVR 119
Db 6976 PMARIAELGLGDDFNQSVVSLPPAVDRDLRALQVLDHDLRLVLPDSTEV 7035
Qy 120 G-----VTOLGG-----AC-----128
Db 7036 APGSVPADVLVSVTRAPGATGEARDALLVEAACARDRLAPAEGRMLQAVLVDGTDGTD 7095
Qy 129 -----SPTWCLITE-----138
Db 7096 GTGGTSGADGVLLVAHLVDSVTWISWPDLAAYRGEEPAVPVGTWSRWQWATSLARLA 7155
Qy 139 -----DTGFDL-----144
Db 7156 TDPVREATAHWEHTLTGAGTLRLDRGRDLOGDAGRISLDLAPHTTEALLTRLPGGVNAS 7215
Qy 145 -----144
Db 7216 VHDVLLTAFANAVAGWRGCGEDPAPVVLDLSEHGRHEEAVPGAELSRTAGWFTALHPV 7275
Qy 145 -----144
Db 7276 RLAPDVTWARLHQDGLRDGLKQVKEQLRSVPGDGLGHLRLHNLPTAGPRLARLPEP 7335
Qy 145 -----144
Db 7336 DFGFNYLGRVTPATGTPPEPMTVTGGLAASRPTAPMAHVELSAVVHEGADGPRRLAEW 7395
Qy 145 -----GVT 147
Db 7396 TYARLVPDHDARRLAEQWFRALAEALVEQADRGAGTGLT 7434

RESULT 9
T27935
hypochemical protein ZK617.1b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000
C:Accession: T27935; T28031
R:White, S.
submitted to the EMBL Data Library, May 1996
A:Reference number: 220442
A:Accession: T27935
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7160 <WIL>
A:Cross-references: EMBL:Z73897; PIDN:CAA98065.1; GSPDB:GNO0022; CESP:ZK617.1b
A:Experimental source: clone ZK617
R:Harris, B.
submitted to the EMBL Data Library, May 1996
A:Reference number: 220458
A:Accession: T28031
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7160 <W12>
A:Cross-references: EMBL:Z73899; PIDN:CAA98082.1; GSPDB:GNO0022; CESP:ZK617.1b
A:Experimental source: clone ZK629
C:Genetics:
A:Gene: CESP:ZK617.1b
A:Map position: 4
A:Introns: 10/3; 61/3; 135/2; 168/3; 256/2; 379/3; 405/2; 463/1; 508/3; 542/3; 574/3; 593/671/1; 3141/3; 3269/1; 6473/3; 7012/3; 7097/1; 7129/3
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

Query Match 29.1%; Score 225; DB 2; Length 7160;
Best Local Similarity 1.2%; Pred. No. 1.2e+02;
Matches 83; Conservative 29; Mismatches 34; Indels 6500; Gaps 26;
Qy 3 GGTLHL-----8
Db 146 GGLYHAIFSDGLDQTYLCOLEIRGSSDAGQVRCNIRNDOGETNANLALNFEPPSPSR 205
Qy 9 -----8

Db 206 QERKRSTASPRSSRGSGSRSPSPKSMKSRCTPKRTLKPREGSPSKLRSRTSTPVNE 265
Qy 9 -----8
Db 266 EVSQSESRSSRTDKMEVDQVSGASKRKPDPGLPPPGDEKKLRAGSPSTRKSPRSKASASP 325
Qy 9 -----8
Db 326 TPRSCKSSAGGAASGTTGASASATSATSGSASSDASRDKYTRPPIVLEASRSOTGRIGG 385
Qy 9 -----8
Db 386 SVLVQWQCHSSTIIWYRDGTLVRNSSEYSQSPNGSIAKLVNKLTEKSGLYKCHAK 445
Qy 9 -----8
Db 446 CDYGEQSSAMVKIEOSDVEEELMKHKOAEDEYKQEEQKQSOTLOAETKKRVARRSKSKS 505
Qy 9 -----ELL 11
Db 506 KSPAPQAKKSTTSRGRQEAASEVHKRSSVRPDPDEESQDDEIPSSGLTIPEERRRELL 565
Qy 12 VAVG-----PDVF-----19
Db 566 GQVGESDDEVSEISLPSFAGCKPRRKTDFSCVLLYLQSPKQDDMFSDTLRKTIT 625
Qy 20 -----19
Db 626 STKNESSTVEEKTARLTKVKKVDGELDFKAMVKLVKVKKEEGGTTEKSGPPLHADSTSS 685
Qy 20 -----19
Db 686 VLSQESRRRGSNAPFAKQGLPEQANPPAQLKKVKSGAGLEKSDSMASLKKDLKKKG 745
Qy 20 -----19
Db 746 IDNDSGAFKVLKVKVKEVKESTISVKEKNGTESGIKTFKWEKERTTLQKYEKTD 805
Qy 20 -----19
Db 806 DGSKKEDKPKKVIAPVSTNKSDDPESTPRHHKEVEKSTSEELKAKVAGQVQKRG 865
Qy 20 -----QAHQ-----23
Db 866 AQKPEEPKLLSQIQLKKVTKKAHDDTNELEGIKLKKVTPKHVADDDSQSESESRGS 925
Qy 24 -----23
Db 926 VFGELRRGRAPRDSADNRRSSIDMRRESVOEILEKTSTPLVPSGASGSAKIV 985
Qy 24 -----23
Db 986 EVPNVTWENETAILTCKVSGSPAPTRFWPKGSREVISGRFKHITDGKEHTVALALLK 1045
Qy 24 -----23
Db 1046 CRSQDEGPYTLTIENVHGTDSADVLLVTSNGLDPRAMLKHRSEAGFQKDGEGGAGG 1105
Qy 24 -----23
Db 1106 GGGEKPMTEAERRQSLFPQKVKVEKWDIPLPEKTVQOQVDKICEMKCTYSRPNKIRWYK 1165
Qy 24 -----EDTERY-----29
Db 1166 DRKEIFSGGLKYKIVIEKNVCTLIINNPEVDDTKYTCCEANGVTHAQLTVLEPPMKYSF 1225
Qy 30 -----VLT-----32
Db 1226 LNPLNTQEIYRTKQAVLTKCVNTPRAPLVWYRGSKAIQEGDPRFIIEKDAVGRCTLTIK 1285
Qy 33 -----32

Db 1286 EVEDDOAEWTARITQDVSFKVQVVEEPRHTFVPMKSKQKNESDLATLETVDNDKDAE 1345
Qy 33 -----
Db 1346 VVWHGDKRIDIDGVKFKVSESSNRKRLLIINGARIEDHGEYKCTTKDDRTMAQLIVDAKN 1405
Qy 33 -----NLNI 36
Db 1406 KFIVALKDETEIEKDDVTLMCQTKDTKTPGIWFRNGKQISSMPGKFETOSRNGHTLKI 1465
Qy 37 G-----
Db 1466 GKIEWNEADYVEIDQAGLRGSCNVTVLEAEKRPILNWKPKKIEAKAGEPCVVKVPOIKG 1525
Qy 38 -----AELLR-----42
Db 1526 TRRGDPRAQILKNKPIDEENRKLVEVIIKDDVAEIVFKNPQLADTGKWALELGNAGTA 1585
Qy 43 -----DPSLGA-----48
Db 1586 LAPPFLVKOKPKPKGPLETKNVTABGLDLVMGCTPDDEGAPVKAVIIEWQEGRSNGWA 1645
Qy 49 -----
Db 1646 KVGETKGTDFKVKDLKEHGEYKFRVKALNECGLSDPLTGESVLAKNPYGVPGPKKNMDAI 1705
Qy 49 -----48
Db 1706 DVDKDHCTLAWPEEEDGAPITGYIIERREKSEKDHQVQOTKPDCCELTDKKVVEDKE 1765
Qy 49 -----48
Db 1766 YLVRKAVNAGPGDPCDHGKPIKWKAKASPEFTGGIGKDLRLKVGETIKYDVPISGEP 1825
Qy 49 -----48
Db 1826 LPECLWVNGKPLKAVGRVWMSRSGKHMKIENAVRADSGKFTITLKNSSGSCDSTATV 1885
Qy 49 -----48
Db 1886 TVVCRPTPKGLDIADVCADGATLSWNPDDGDDPLTGYVIEAQDMNDKGVIEVGKV 1945
Qy 49 -----48
Db 1946 DPNTTLKVNGLRNKGNVFRVKAVNNEGESEPLSADQYQIKDPWDEPGKPGRPEITDF 2005
Qy 49 -----48
Db 2006 DADRIDIAWEPHPKDGGAPEEYIVEVRDPDTKEWKEVRVPDNTNASISGLKEGEYQFR 2065
Qy 49 -----48
Db 2066 VRVNVKAGPQGPSEPEKQLAKPKFIPAWLKHDLNLSITVKAGATVRWEVKIGGEPIPEV 2125
Qy 49 -----48
Db 2126 KWFKNQOLENGIQLTIDTRKNEHTILCIPSAMRSDVGEYRLTVKNSHGADEEKANITVL 2185
Qy 49 -----48
Db 2186 DRPSKPNGLVSDVFEDNLNLSWKPPDDGGEPYIEYEVEKLDATATGRWVPCAKVKDTK 2245
Qy 49 -----48
Db 2246 AHIDGLKKGQTYQFRVKAVNKEGASDALSTDKOTKAKNPYDEPKGTGTPDVVDWADRVS 2305
Qy 49 -----QFRV---52
Db 2306 LEWEPKSDGGAITYQVIEKKKGHRDWCQCGKVGSDQTNAEITLGLKEGEYQFRVAV 2365
Qy 53 -----52
Db 2366 NKAGGEASDPSRKVAVKPNLKPWIDREAMKTITIKVGNDEPDPVVRGPEPPKKEWIF 2425

Qy 53 -----HLVQMVL-----60
Db 2426 NEKPVDDQKIRIESEDYKTRFVLRGATRKHAGLYTLTATNASGDKHSVEVIVLGPSSP 2485
Qy 61 -----60
Db 2486 LGPLEVSNVYEDRADLEWKVPEDDGGAPIDHYEIEKMDLATGRWVPCGRSETTKTTVPNL 2545
Qy 61 -----60
Db 2546 QPGHEYKFRVAVNKEGESDPLTTNTAILAKNPYEVPGVKDPBELVDWDKDHVDLAWNAP 2605
Qy 61 -----60
Db 2606 DGGAPIEAFVIEKKDKNGRWEALVVPDQKTATVPNLKEGEYQFRISARNKAGTGDP 2665
Qy 61 -----60
Db 2666 SDPSDRVVAKPNLAPRIHREDLSDTTVKVGATLKFIVHIDGEPAPDVTWMSFNKGIGES 2725
Qy 61 -----TEPEG-----65
Db 2726 KQAIENEPVISRPALPKALRKOSGKYITATNINGTDSVTINIKVKSCKTPKPGPIEVD 2785
Qy 66 -----65
Db 2786 VFERATLDWKPPEDDGGEPBIEFYIEKMTKOGIWWPCGRSGDTHFTVDSLNLKGDHYKF 2845
Qy 66 -----65
Db 2846 RVKAVNSEGSDPLETETDILAKNPPDRPDRPORPEPTDWDSDHVDLKWDPPLSDGAPI 2905
Qy 66 -----65
Db 2906 EBYQIEKRTKYGRWEPAITVPGQTTATVPDLTPNEEYEFRRVAVNKGSPSPDASKAV 2965
Qy 66 -----65
Db 2966 IAKPNLKPDIRDALKNLTIKAGOSISFDPVSGSPAFTVTWHWPDNREINRGVRKLD 3025
Qy 66 -----65
Db 3026 NPEYQSKLVVQMERDGSCTFTIKAVNAGEDEATVKINVIDKPTSPNGPLDVSVDHGDH 3085
Qy 66 -----65
Db 3086 VTLNWRAPDDGGIPIENYVIEKYDTASGRWVPAKAVAGDKTTAVVDGLIPGHEYKFRVA 3145
Qy 66 -----APNIT-----70
Db 3146 AVNAEGSDPLETFTGTTLAKDPDPKPKTNAPEITDWDKDHVDLEWKPPANDGGAPIEEY 3205
Qy 71 -----70
Db 3206 VVEMKDEFSFPFNDVAHVAPAGOTNATVGNLKEGKYEFIRAKNAGLGDPSDSASAVAK 3265
Qy 71 -----70
Db 3266 ARNVPPVIDRNSIQEIKVKAGODFSLNIIVSGEPTTITWTEGTPVESDDRMKLNEDG 3325
Qy 71 -----70
Db 3326 KTKFHVKRALRSDTGTGTYYIIKAENENGTDTAEVKVTYVLDHPSSPRGPLDVNTNIVKDCDLA 3385
Qy 71 -----70
Db 3386 WKEPEDDGAIEISHYVIEKQDAATGRWTACGSKOTNFHVDLDTQGHYKFRVKAVNRHG 3445
Qy 71 -----70
Db 3446 DSDPLEAREAILIAKOPDFDRADKPGTPEIVDWDKDHADLKWTTPPADGGAPIEGYLVEMR 3505

Qy 71 ----- 70
 Db 3506 PSGDWPAVTVGAGELTATVDGLKPGQTYQFRVKALNKAGESTPSDPRTWVAKPRHLAP 3565
 Qy 71 ----- 70
 Db 3566 KINRDMFVAQRVKAGQTLNFDVNVVEGEPAPKIEWFLNGSLSSGGNTHIDNNTDNNTKLT 3625
 Qy 71 ----- 70
 Db 3626 TKSTARADSGKYIVATNESGKDEHEVDNLDIFCAPEGPLRHKDIITKESVVLKWDDEPL 3685
 Qy 71 -----ANLT 74
 Db 3686 DDGSPITNVVVEKQEDGGRWVQCGETSOTSLKVNKLSEGHYKFRVKA VNRQGTSAFLT 3745
 Qy 75 S----- 75
 Db 3746 SDHAIVAKNPPDEPDAPTDVTPVDWDKDHVDLEWKPANDGGAPIDAVIVEKKDKFGDWV 3805
 Qy 76 ----- 75
 Db 3806 ECARVDGKTKATADNLTPGETYQFRVKAVNKA PGKPSDPTGNVAKPRRMAPKLNLAG 3865
 Qy 76 ----- 75
 Db 3866 LLDLRKAGTPIKLDIAFEGEPAPVAKWKANDATIDTGARADVNTPTSSAIHIFSARVG 3925
 Qy 76 -----SLLSVCQ----- 82
 Db 3926 DTGVYKIIIVENEHKGDTAOCNVTLVDVPGTPEGPLKIDEIHKEGCTLNWKPPDPTDNGGTDV 3985
 Qy 83 ----- 82
 Db 3986 LHYIVEKMDTSRGTWQEVGTFDCTAKVNKLVPKGYAFRVKAVNLOGESKPLEAEPII 4045
 Qy 83 -----WSQTIN----- 88
 Db 4046 AKNQFDVDPVDPKPEVTDWKDRIDIKMNPNTANNGGAPVTGTVIVEKKEGSAINTWAGKT 4105
 Qy 89 -----PEDDTDP----- 95
 Db 4106 PGTTSADNLKPGVEFEFRVIAVNAAGPSDPTDPTDQITKARYLKPILTASRKIKKA 4165
 Qy 96 ----- 95
 Db 4166 GFTHNLEVDFIGAPOPTATWTVGDSGAALAPELLVDAKSSTTSIFPPSAKRADSGNYKLK 4225
 Qy 96 ----- 95
 Db 4226 VKNELGEDEAIFEVIVODRPSAPEGPLEVSDVTKDSCVLNWKPPKDDGGAETSNVYVEKR 4285
 Qy 96 ----- 95
 Db 4286 DTKNTWVPVSARFVTCSTIVPKLTGEGHEYEFVRVMAENTFGRSDSLNTDEPVLAKDPFGT 4345
 Qy 96 ----- 95
 Db 4346 PGKPGRPEIVTDNDHIDIKWDPDRNGGSPVDHYDIERKDAKTGRWIKVNTSPVOGTAF 4405
 Qy 96 ----- 95
 Db 4406 SDTRVQGHYTEYRVRVAVNAKAGQPSDSSAAATAKPMHEAPKFDLDLDCKEFRVKA GEP 4465
 Qy 96 ----- 95
 Db 4466 LVITIFTASQPDI SWTEGGKPLAGVETDTSOTKLVIPTSTRSDSGPVKIKAVNPYGE 4525
 Qy 96 ----- 95
 Db 4526 AEANIKITVIDKPGAPENITYPAVSRHTCTLNWDAPKDDGGAIEAGYKIEYQEVGSOIWD 4585
 Qy 96 ----- 95

Db 4586 KVPGLISGTAYTVRGLEHGQQYFRIRAEANAVGLSDYCOGVPVVIKDPFPPGAPSTPEI 4645
 Qy 96 ----- 95
 Db 4646 TGYDTNOVSLAMNPPRDDGGSPILGYVVERFEKRGGGDWAPVKMPVMVKGTECIVPGLHEN 4705
 Qy 96 -----GH----- 97
 Db 4706 ETYQFRVRAVNAAGHGSPSNGSEPVTCRPYVEKPGADAPRVGKITKNSAELTNRRLRD 4765
 Qy 98 ----- 97
 Db 4766 GGAPIDGYIVEKKKLGDNWTRCNDKPVRODTAFEVKNLGEKEEYEFVRVIAVNSAGEPS 4825
 Qy 98 --ADLVI----- 102
 Db 4826 KPSDLVILIESQGRPIPDINNLDITVRAGETIQIRIPYAGGNPKPIIDLFGNSPIPEN 4885
 Qy 103 ----- 102
 Db 4886 ERTVVDVNPGEIVITTTGSKSDAGPYKISATNKYKDKTCKLNVFVLDAPGKPTGPIRAT 4945
 Qy 103 ----- 102
 Db 4946 DIQADAMTLSWRPPKNGDAITNYVVEKRTPGGDMWTVGHVPGTTLVRNLDANTPYEF 5005
 Qy 103 ----- 102
 Db 5006 RVRAENQYVGCEPLETDDAIVAKNPPDTPGAPQOPEAVETSEEAITLQWTRPTSDGGAPI 5065
 Qy 103 ----- 102
 Db 5066 QGYVIEKREVGSTWTKAAGNILDTKHRVTGLTPKKTTEFRVAAVNAAGOGESVNSVP 5125
 Qy 103 ----- 102
 Db 5126 ITADNAPTRPKINMGMLTRDILAYAGERAKILVPFAASPAPKVTFSGENKISPTDPRVK 5185
 Qy 103 ----- 102
 Db 5186 VEYSDFLATLTIEKSELTDGGLYFVELENSOGSDSASIRLKVVDPKASPQHIRVEDIAPD 5245
 Qy 103 ----- 102
 Db 5246 CCTLYWMPSSDGGSPITNYIVEKLDLRHSDGKWEKVSFVRNLNNTVGGLIKONRYRFR 5305
 Qy 103 -----YITRFDLELP----- 112
 Db 5306 VRAETQYGVSEPCELADVVAKYQFEVNPNOPEAPTVRDKDSTWAELEWDRDRDGGSKIIG 5365
 Qy 113 ----- 112
 Db 5366 YQVQYRDTSSGRWINAKMDLSEOCHARVTGLRONGEFEFRIIKNNAAGFSKPSPPSERCQ 5425
 Qy 113 -----DGNROVRG----- 120
 Db 5426 LKSREPGPGPIHVGAKSIGRNHCTITWMAPLEDGGSKIICGVNVEIREYVSTLWTVASDY 5485
 Qy 121 ----- 120
 Db 5486 NVREPEFTVDKLRBENDYEFVRWAINAAGKGIPLSGPIKIOESGSRPQIVVAKPETA 5545
 Qy 121 ----- 120
 Db 5546 QPYNNRAVFTCEAVGRPEPTARWLNRNGRELPESSRYRFEASDGVYKFTIKVWDIDAGEY 5605
 Qy 121 ----- 120
 Db 5606 TVEVSNPYGSDTATANLVVQAPPVIEKDVNPNTILPSGDLVRLKIYFSCTAPFRHSLVLR 5665
 Qy 121 ----- 120

Db	5666	BEIDMDHPTIRIVFDDHILITIPALS	VREAGRYEYTVSNDGSEATTGFWLNV	TGLPEAP	5725	
Qy	121	-----	-----	-----	120	
Db	5726	OGPLHISNIGPSTATLSWRPPVTDGGS	KITSYVWEKRDLSKDEWTVTSNVKDM	NYIVTG	5785	
Qy	121	-----	-----	-----	126	
Db	5786	LFENHEYFRVSAQNGENGICAPLVSEHP	IIARLPDPPTSPNLLEIVQVGDDYVTL	WSQR	5845	
Qy	127	-----	-----	-----	126	
Db	5846	PLSDGGRLRGYIVEKQBEHDEWF	CNQNPPNNYVNPILDGRKYRVFVAV	NDAGL	5905	
Qy	127	-----	-----	-----	126	
Db	5906	SDLAELOTLFOASGSGEPKIVSPL	SDLNBEVGRCVTFECEISGSPREYR	WFKGCKEL	5965	
Qy	127	-----	-----	-----	126	
Db	5966	VDTSKYTLINKGDKQVLIINDLTS	DDADEYTCRATNSSGSTRANLR	IKTKPRVFIPPK	6025	
Qy	127	-----	-----	-----	126	
Db	6026	YHGGVEAQKGETIELKIPKAYPO	GEARWTKDGEKIENNSKFSITDDK	FATLRISNASR	6085	
Qy	127	-----	-----	-----	126	
Db	6086	EDYGEYRVVSVNSGSDGTVNV	TVADVPEPPRFPIIENILDEAVIL	SWKPPALDGGSLV	6145	
Qy	127	-----	-----	-----	131	
Db	6146	TNYTIEKREAMGGSWSPCAKSR	YTYTTIEGLRAGQYEFRIIAENKH	GQSKCEPTAPVL	6205	
Qy	132	-----	-----	-----	131	
Db	6206	IPGDERKRRGYDVDEQGI	VRGKGTSSYNDNYVFDIWKQY	POPVEIKHDVLDHYDI	6265	
Qy	132	-----	-----	-----	131	
Db	6266	HEELGTGAFGVVHRVTERAT	GNNFAAKFVMTPHESDKETVRKE	IQTMSVLRHPTLVN	LHD	6325
Qy	132	-----	-----	-----	131	
Db	6326	AFEDDNEMVIYEFMSGGEL	FEKVADEHNKXSEAVEYMRQV	CKGLCHMENVVH	LDL	6385
Qy	132	-----	-----	-----	133	
Db	6386	KPENIMFTTKRSNELKJID	FGLTAHDPKQSVKVTGTGA	FAAPEVAEGKPVGY	YTDMS	6445
Qy	134	-----	-----	-----	133	
Db	6446	VGVLSYILLGLSPFGEND	DETLRNKVKCDWNDDSA	FSISGDKDFIRKLLAD	PNT	6505
Qy	134	-----	-----	-----	133	
Db	6506	RMTIHALEHPLTPGNAP	RDSQIPSSRYTKIRDSIKTKY	DANWPEPLPLGR	ISNYSSL	6565
Qy	134	-----	-----	-----	133	
Db	6566	RKHPQEYSIRDAFWR	SEAQPRFIVKPYGTEVG	EGSANCYFCRVIAS	PPVTHKDDR	6625
Qy	134	-----	-----	-----	133	
Db	6626	ELKQSVKMKRYNGNDY	GLTINRVKDDKGEYTVRAKNS	YGTKEIVFLNTRH	SEPLKF	6685
Qy	134	-----	-----	-----	137	
Db	6686	EPLPMKKAPSPRVEE	KERRSAPFTFHLNR	LIQKHOCKLTC	SQGNPNTIEMWK	6745
Qy	138	-----	-----	-----	148	
Db	6746	DGHPVDEDRVQVSR	FGVCSLEIFNARV	DDAGTYTVTATND	LGDV	6791

RESULT 10
T31677

bacitracin synthetase 1 - Bacillus licheniformis
C:Species: Bacillus licheniformis
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Nov-2000
C:Accession: T31677
R:Konz, D.; Klena, A.; Schorgendorfer, K.; Marahiel, M.A.
Chem. Biol. 4, 927-937, 1997
A:Title: The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: mo
A:Reference number: Z21058; MUID:98089193; PMID:9427658
A:Accession: T31677
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5255 <CON>
A:Cross-references: EMBL:AF007865; NID:g4464275; PID:g2982194; PIDN:AAC06346.1
C:Genetics:

C:Gene: bacc
C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology
C:Keywords: carrier protein
F:83-524/Domain: acetate-CoA ligase homology <ACLI>
F:542-610/Domain: acyl carrier protein homology <ACLI>
F:1134-1561/Domain: acetate-CoA ligase homology <ACLI>
F:1583-1651/Domain: acyl carrier protein homology <ACP2>
F:1711-2601/Domain: acetate-CoA ligase homology <ACL3>
F:2619-2687/Domain: acyl carrier protein homology <ACP3>
F:3195-3642/Domain: acetate-CoA ligase homology <ACL4>
F:3662-3729/Domain: acyl carrier protein homology <ACP4>
F:4712-5151/Domain: acetate-CoA ligase homology <ACL5>
F:5169-5237/Domain: acyl carrier protein homology <ACP5>

Query Match 29.0%; Score 224; DB 2: Length 5255;
Best Local Similarity 1.6%; Pred. No. 47;
Matches 77; Conservative 34; Mismatches 36; Indels 4568; Gaps 24;

Qy	2	AGG	-----	4
Db	110	AGGAYLPIDETPKRIAPMLSDTKAAVLLTQGAADIGDCEADIVOLDREASDGSFKEP	169	
Qy	5	-----	-----	4
Db	170	LSSVNSDGTAYIYTGSGTGPKVITPHYSVIRVVQNTNYDITEDNVLQLSNYSFD	229	
Qy	5	-----	-----	4
Db	230	GSVDFIPGALLNGASLVMIEKEALLNINRLGSAINEEKYSVMFITTALFNMIADIHVDC	289	
Qy	5	-----ILHL-----	-----	8
Db	290	SNLRKILFGGERASIPHRKVLNHVGDRKLIHVYGPTESTVYATYFINEIDDEAETIPI	349	
Qy	9	-----	-----	8
Db	350	GSPLANTSVLIMDEAGKLVIPGVGELCTAGDGLSKGYLNREBELTAEKFIHPHFPIGERL	409	
Qy	9	-----	-----	8
Db	410	YKTGDLAKWLPDGNIEFIGRIDHQVKIRGPFRIELGESRLEMHEDINETIVTVREDES	469	
Qy	9	-----	-----	8
Db	470	RPYICAVITANREISLDELKGLGKLPPEYMIPAYFVKLDPKLPKNGKVDKALPEPDR	529	
Qy	9	-----ELLVAVGPDVFQA-----	-----	21
Db	530	TAGAENEYAPRNETEKLAAVQDVLHVEKAGIHDHFAQMGHSLHAMELIAKIKEMN	589	
Qy	22	-----HOED-----	-----	25
Db	590	VEIPLHQLFKLATIKELSAFIEANHOEDKGDTLVTRAADPENIHEIFPLTGIQLAYLVGR	649	
Qy	26	-----	-----TER--	28

Db 650 DETFEIGGVATNLTVFEADVDLNRFLQTLQKLIDRHPILRTIVFENGSTQKILEATQRYT 709
 QY 29 -----
 Db 710 IETQDLRGFTBEEINVRILEQREKMTSKIIDPSVMPLFELKTFMPLGEEKYFFLNVDPLI 769
 QY 29 -----
 Db 770 CDDSSMKRLIREFKQLYENPGLQSPLEYSFRDYVLASINFKQTSRYOKDQOQYWLKDLH 829
 QY 37 -----
 Db 830 FPSAPELPLKSDPAHVAKPSFKFSTFLDGHWTWNLKKKARHHHLTPTSVLCAAYAYILA 889
 QY 37 -----
 Db 890 YMSRQNHFAINTVFNRIPFHPDVKMIGDFTSLMDDIHABENMSFWRFALNVQDTLL 949
 QY 37 -----
 Db 950 EALEHRHYDGVDIRNIKAKNGKNKAVMPIVFTSVLSENPDSDSLVDPNHIFSTR 1009
 QY 43 -----
 Db 1010 TSQVVIDNOVYEINGGLYITWDYEQIPEHEVIESMFDQYIAVIQKAVSGEDVSTIQMNE 1069
 QY 43 -----
 Db 1070 KSRQMSAYNDTDQSFDAKPLHELFTGVQVKGHPDRMALKHDEVMTYQELDEKSNQVAF 1129
 QY 43 -----
 Db 1130 LIGKVEKGYIGVIGKRSGLTIVNLLAVLKTGAAYIPLDPDYPERKAYIOSKSNCKFF 1189
 QY 45 -----
 Db 1190 ISHDVYDKEHIERFSKAPVDRKVDLDDMAYVIFTSGSTGPKGVQITHVPQRTILDINE 1249
 QY 45 -----
 Db 1250 KFNVTQDNIMIGISSLCFSLVYDVFGLSSGASLVIIDDQDVFSLKETAERITWN 1309
 QY 53 -----
 Db 1310 SVPAINGMTADVDPNENLHRLILLSGDWIPLQLPATIKKTFKNAEVISLGGATEGI 1369
 QY 61 -----
 Db 1370 WSIYPIQKVEEDWKSIPYKPLANOKIYVLNQNKOLCPVGEVELYIGGAGVASGYIHD 1429
 QY 61 -----
 Db 1430 QEKTEHSFIHQELGYIYKTGDYGLVEDGYVEFLGRKDSQVKIRGYRVMGEIENTLVS 1489
 QY 61 -----
 Db 1490 HOBITKASVIDYTSPOGINKLYAFVVAENAIQSOLDVKEFLQKTLPOYMIIPAKFVQIEEP 1549
 QY 72 -----
 Db 1550 LTVNGKVDKRTLHDLAEQHTADEGQGRMLPENETQAMLLEIWKDIFGLDSINLDVSY 1609
 QY 72 -----
 Db 1610 EIGGDSLKAIISIINKNRNVEMPISEIPKNDITIALDHYLKNREDSMEHPQKAREK 1669
 QY 72 -----
 Db 1670 EYPTSPAQQMYMLSLENERGAYHIPMALLVEGRINAMOLENALKTLQRIEILRTGF 1729
 QY 72 -----
 Db 1730 LRAEVVKNREHLLMFDMHHIISDGVSTDIIFIOELGALYBECKSLKPFHIQYKDYAEWEN 2869

Db 1730 EIQNNELIQIYENVDRFLREYECLDASITDOHALMEITSRYCKESIKPPDLSRPLMRK 1789
 QY 76 -----
 Db 1790 LJKIDDIRHILVINFFHHIISDGVSQGILMNEILELYSNVPLPEVNVQYKDYVENHTPNQ 1849
 QY 76 -----
 Db 1850 SAMKKQEAAYWLDVYRDIPSKLDPPDYDKRHHIDTPEGSVFLEMERELSDHIRKLAKHN 1909
 QY 76 -----
 Db 1910 GTTLYTVMLSAYYVLLNKYTNQTDIVVGTAAGRSLHPLQDQVGFVNTIALRNEVTSY 1969
 QY 76 -----
 Db 1970 SFKEFLOOTKERTTAAPDNSEYFPDDLIRKLVGVRNSRNPPLFDTMFVLEDAARMFTKQG 2029
 QY 76 -----
 Db 2030 DVKLSPIIFELDNAKPDMI FNVLDPEOKIVLNIEYSTSLFKDETIOKIAEDYPRILEVS 2089
 QY 76 -----
 Db 2090 ENLDVALHQIDMISROEKRTLLESFNHTKAYPKGKAIHOLFEEQAKRI PDHTAVVPEDQ 2149
 QY 76 -----
 Db 2150 KLYRQLNEKANQVARELLREKGVKPDTLVGIMMERSSDMTAAIILGVLKAGAYLPIDPEY 2209
 QY 76 -----
 Db 2210 PPERMYMAFDSEVKVVIISDVPLAEELTAESIILHMDDERIAGDRSDIDNVNQSGLA 2269
 QY 76 -----
 Db 2270 YVIYTSSTGKPKGMIEHOSLNLCSMHOSCFEVGQNDNSSIYASISFDAFVWELPPYI 2329
 QY 86 -----
 Db 2330 TAGATVHLVNOETRLDVEKLNRYPHDHHITISFLPTVCEQFTALDNHSLRLLTGGDKL 2389
 QY 86 -----
 Db 2390 NVFKEKSVQIVNNYGPTENTVWATSPIDKSHQNIPIGKPIDNVKVLNKDQLQCLGA 2449
 QY 86 -----
 Db 2450 SGELCIAGEGLARGVNRPELTREKFIGNPFVPGMYRTGDLAKMLPDGNIQFLGRVDQ 2509
 QY 86 -----
 Db 2510 QVKIRGYRIEPEIENLLKYKEIEEAAVIAREDCDHPYLCAYTVTVKVEPEKIRAPL 2569
 QY 86 -----
 Db 2570 KKSLPDYMIPOYFVQLDGLPLTVNGKVDKKSLLPVPERSVTMDRRYEA PRDQMEKLVSIW 2629
 QY 92 -----
 Db 2630 EEALGINKIGINSHFFEAGGHSKAAALVSTIHKELNVKPLRQIFETPTIKGLRDISVR 2689
 QY 92 -----
 Db 2690 RRKCFYIDRKTEEPYRSLSSAQKRLYILSOTGSHVAYNMPFAMTLEGDFDIRRENTLK 2749
 QY 92 -----
 Db 2750 NMVKHESFRTSFVMIDGEMVQOIEKEIDFQVAYSDICKESAEEKIKSIFRPFHLEKAPL 2809
 QY 92 -----
 Db 2810 LRAEVVKNREHLLMFDMHHIISDGVSTDIIFIOELGALYBECKSLKPFHIQYKDYAEWEN 2869

```
Qy 92 ----- 91
Db 2870 SHARSELKROEYMLTKYKGI PVLDPIDHKRLTKSSEGDVTAAIESETRKLQHM 2929
Qy 92 ----- 91
Db 2930 AKENGVTMYMLLAGYTALLSKYTQEDII VGTPAAGRNHEDIQHILGMFVNTLAIRNH 2989
Qy 92 ----- 91
Db 2990 EGKKTFRDYLOEVKENTLQAYENQDY PPEELVEKVINIKRDMARNPLFDTMLVYHNTDKP 3049
Qy 92 ----- 91
Db 3050 FEAELRSRLVEIKRGISKFDITWASEAADGLRLEVEYSTTLFNKERMERLSEHLISLL 3109
Qy 92 ----- 91
Db 3110 EQAADHPDIAINQIDVLTGERHRVLDNFRTDGVFCKEMTIPELFEKQAEKTPDHPAVA 3169
Qy 92 ----- 91
Db 3170 FGDETISVRELNERANSIAFTLRQKGVCPDVIAGILTERTSIEMIVGIMGILKAGAYLPI 3229
Qy 92 ----- 96
Db 3230 DPAYQERISIVKDSVSVLCAAGDVPGEAYTGDIIRIDQTGONDHVENLKHDIKPOH 3289
Qy 97 ----- 96
Db 3290 LAYVIYTSGSTGPKPGVMI EHHSVNNLVHGLNERIYOHLDALHVALVAPYIFDASVKQI 3349
Qy 97 ----- 110
Db 3350 FAALLFGHTLCIVRETAMDAMSL EYYSKNNINVSMDTPAHLNMLAYVDKTELEFDVKE 3409
Qy 111 ----- 110
Db 3410 LIVGGDALTPDIVGGLPHKPNLS CNITNVYPTCCVDAASHQIESGKVPQTPSIPIGR 3469
Qy 111 ----- 110
Db 3470 PLLNTSIYIVDKELRPLVPVGIAGELCIAGEGVARGVYVNRPELTAEPVDFHFPFGKMYK 3529
Qy 111 ----- 114
Db 3530 TGDAMWLPDQIEFLGRADHQVKIRGYRIELGEVEQQLLTHEKIKEAAVIAKQDQNGS 3589
Qy 115 ----- 114
Db 3590 YLCAYIASDKELPAADVROFLEREMPDYMI PSYFVKLDRLPRTPSGKVDRSALPEADGNV 3649
Qy 115 ----- 114
Db 3650 NVMEGTGYPDRNEIERKLVQVREILGABD IGI SHHFFAAGGDSIKALQIVSR LAKMNL 3709
Qy 115 ----- 124
Db 3710 KLEMKALFANPKIKOLSRFITEETRHKG NKVPVGETELLPIQKRYFANNKEELDHFNOS 3769
Qy 125 ----- 124
Db 3770 FMLFRKGDYDENIVRTAFN KILEQHDALRM IYEEKDGI IQVNRGYRENFLDLDVYDVVRG 3829
Qy 125 ----- 124
Db 3830 FDSQEEKVELATG IOKSSIRKGLVHLGI FRADEGDHLLTAIHLVVDGVSWRILFED 3889
Qy 125 ----- 124
Db 3890 FETLYLOALKEPLD IGYKTDTSYQEFARQL KYAQSRRLKEREYQWQKALEADVFPFAE 3949
```

```
Qy 125 ----- 124
Db 3950 KLERDTPHSATLSIRIGPDVTAKLLRNAPKAYNTEINDILLTALIAAVRDITGENKLKV 4009
Qy 125 ----- 124
Db 4010 MMEGHGRELDIGVDITRTIGWFTTVVP FIDLGESEKISQNIKMVKEALRKIPNKGIGY 4069
Qy 125 ----- 124
Db 4070 GVLKYMTEELQKIQTQAPLPSFN YFGEMNDMNRKVFSQSPFSGESIGGKIVRHCAIEMN 4129
Qy 125 ----- 131
Db 4130 AISLNGELTIYTTFNQDQYQSTI EQLNQSPKENLEKIVDHCVDKESDWTSPDYGDVSL 4189
Qy 132 ----- 131
Db 4190 GLELELIKDKYSAFOIEKIYPLANM QKMLFHNAMDOTSGAYFOOIVIKLGRVHPDIL 4249
Qy 132 ----- 131
Db 4250 EESFHEIVKREHLRASFEYBITAE PRQIIARDRKTPTTSIDLTGENRTRQHRFIETYLK 4309
Qy 132 ----- 133
Db 4310 EDOEKGFDSLSEALMRVCLIKMSD ESRYRLIWSHHILLDGCWLGIVLSLFSLYGKIMKG 4369
Qy 134 ----- 134
Db 4370 ESRLKEPKPYGDIKWLEKQDQBEA VYWKDYLGYESRSELPAPNRGATSEYCGKEK 4429
Qy 135 ----- 134
Db 4430 VISFSKELTTKITRIAKQHHVTINT VLOGIWGMILAKYKNTDEVVFGTVVSGREAPVGOI 4489
Qy 135 ----- 137
Db 4490 EBMVGLFIHTIPTRISFEGARSF KEVLKKTQAESNESRYSYMNILSEIQVLSEMKRELIT 4549
Qy 138 E ----- 144
Db 4550 HYMAFONYAFDBELFRSOSGETFEGE VHGKERTNYNPNLTGVLEDBQLKLKLTFNENV 4609
Qy 145 ----- 144
Db 4610 YDNTIITLEXHIIITVAEQVAEDETQ TLRDINLVSKSEQHRLHTFNTDGTGYPKDKPLH 4669
Qy 145 ----- 144
Db 4670 ELFEQAMKTPDHTALVFGAQRMTYR ELNEKANQATARLLREKIGRGSIAAIIADRSFEM 4729
Qy 145 ----- 144
Db 4730 IIGIIGILKAGGYLPIDPETPRDRID YMLKNVSGAALLVTTSLLKPPDIKTVDLCSDEL 4789
Qy 145 ----- 148
Db 4790 HLLSEENLPRVNRSDTAYIIVYTS GSGTGPKGVI 4824
```

RESULT 11

T17428

FK506 polyketide synthase - *Streptomyces* sp. (strain MA6548)C:Species: *Streptomyces* sp.

A:Variety: strain MA6548

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000

C:Accession: T17428

R:Motamedi, H.; Shafiee, A.

Eur. J. Biochem. 256, 528-534, 1998

A:Title: The biosynthetic gene cluster for the macrolactone ring of the immunosupprea

A:Reference number: Z18779; MUID:98451508; PMID:9780228

A:Accession: T17428

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7576 <MOT>
A:Cross-references: EMBL:AF082100; NID:g3798623; PID:g3798624; PIDN:AAC68815.1
A:Experimental source: strain MA6548
C:Genetics:
A:Gene: fkbB
C:Function:
A:Description: involved in synthesis of the backbone of the immunosuppressant FK506 poly
C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C:Keywords: carrier protein
F:54-500/Domain: acetate-CoA ligase homology <ACL>
F:1095-1166/Domain: acyl carrier protein homology <ACPI>
F:1204-1599/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F:1682-1953/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F:2680-2751/Domain: acyl carrier protein homology <ACP2>
F:2804-3198/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F:3295-3569/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F:4320-4391/Domain: acyl carrier protein homology <ACP3>
F:4435-4810/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F:5903-5974/Domain: acyl carrier protein homology <ACP4>
F:6018-6412/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F:6513-6785/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F:7450-7521/Domain: acyl carrier protein homology <ACP5>

Query Match 29.0%; Score 224; DB 2; Length 7576;
Best Local Similarity 1.3%; Pred. No. 1.7e+02;
Matches 95; Conservative 17; Mismatches 34; Indels 7287; Gaps 28;

QY 1 AAGGI----- 5
DB 51 AASGVGRDVRLLGNRIETIEVLAARAAAVAVPLNPDAGDAELAHFLTDSGATVLV 110
QY 6 LHELLVAVGPD----- 17
DB 111 TDLLHDQVRRVGTATVVLAGRREPGCVAYEDLAGTEPPRPDDGLDEPAMWLTSG 170
QY 18 -VFOAHO-----EDTE----- 27
DB 171 TTGRPGVVSAGRSGLWSATYCDVPSWELTDEDELLWAPLPHSLGHLLCLLAVLVGAS 230
QY 28 -----RYVL----- 31
DB 231 ARILGFGVARDVLDALAEHPCTVLVGVPTMYRYLLGAVSGEPOTRALRVALVAGSTSPAS 290
QY 32 ----- 31
DB 291 LTAGPEAAFGVPLLDYGTCTTGTSLTANTLTDTTRVPGSCGLPVPGLSLRFVDPVTGADV 350
QY 32 ----- 31
DB 351 RRGEGELWASGPSMLGLGHAQTEATAQVLADGWYRTGDLARQAETGHVITGRVKELII 410
QY 32 ----- 31
DB 411 RGENIHPREIESVAQEPGVGRDAAAAGRPVLPVLYVVPVGPVVPADAILAECCR 470
QY 32 ----- 31
DB 471 QLAYFKVPDEILHVTVPRTASGVKVRGELAGLPARLVGTSGGEAALCELWERRLEPDT 530
QY 32 -----TNLNIGA 38
DB 531 VSQVQAVVTRAVGIGPAEVPDADQALRWDEALVRQAAEPGPFVLVDTDITDITDANNGD 590
QY 39 E----- 39
DB 591 EADELVAAAAALGEPQVALRSGVAVPRLVRADTTPLPASQWALRPPASGTLRLDAVAPS 650
QY 40 ----- 39
DB 651 DAPPRPLAAGEVRVDVRAAGLNFRDLVIALGTYPGDGANGGEAAGIIVTEVPGVDDLVPG 710

QY 40 ----- 39
DB 711 AHVFLVQDAFRFRSVVADRRRLVARIPRGWSPFVAASVPIVFATAWYGLVDAGQVRPGQKV 770
QY 40 ----- 39
DB 771 LIHAATGGVGSAAROIARHMGADVYATASPAKQHLLLADGFDADHVANSRTTAPADAFPP 830
QY 40 ----- 39
DB 831 VDVVLSLTGFLDASVELLAPGGRFTIEMGKTDIRHDVQQPFDLSDVDPARLEILL 890
QY 40 ----- 39
DB 891 ELFRDELRLPLVPRWDIRRDARDFGMMOARNTGKTVLTVPRPLGPDAPVLVTGEGADA 950
QY 40 -----LLRDP----- 44
DB 951 VAGLLRDETGRTHVSSSSAAEAEEGEGVNSTLAHVLLGDPAGADAYARSASEGLP 1010
QY 45 -----SLGAO----- 49
DB 1011 IIVVHGLPGIEQAALTGELIEAAIEGGSYVTVRVGSAGARAEAMAGATPILLSGLAGPT 1070
QY 50 ----- 49
DB 1071 GRAEPGAAGPEWMSRLAAADADRENVLDDLVRDSVATVGLPGAERCAPDRTFRENGLDS 1130
QY 50 -----FRVHLVQVVI----- 59
DB 1131 LTTVEFANTVAARTGLRVPASTAFDHTPRSFAAHLGASVAAAPAAKALGSGEPVAIVGM 1190
QY 60 -----LTEPEG----- 65
DB 1191 ACLPQGVTSPEGLWRLVASGTDAITFPVDRGWVDVLPDPDPAVGRSTTRHGGFLTE 1250
QY 66 -----APN----- 68
DB 1251 ATGFDAAFGCIISNEALAMDPOORLVLETSWEAFHAGIVPDTLRESDTGVFMGAPHOGY 1310
QY 69 -----ITANLTS----- 75
DB 1311 GACRDLGLGVTTATQTSVLGRLSYFYVLOGPAAVTVDTACSSSLVALHQAALRSCEGS 1370
QY 76 ----- 75
DB 1371 LALAGGVTVNATPGSPVFPFQRGLSPDGRCKAFADAADGTGPAEGVGLVWVERQSDAER 1430
QY 76 ----- 75
DB 1431 NGHTVLAVVRGSVAVNODGASNGLSAPNGVAQORVIRQALANAGLSGADVDAVEAHGTGTV 1490
QY 76 ----- 75
DB 1491 LGDPIEAQAVLASYGQEREVPLLLGSKVSNVGHGTAAAGVAGVIKVMAMRHGLPKTLH 1550
QY 76 ----- 75
DB 1551 VDEPSSHVDWSAGAVEVTVETRPWPESGGARRAGVSSFGVSGTNAHVILEGVAESSVRDG 1610
QY 76 ----- 75
DB 1611 GSSGLVPLVPSARTESSLALQVERLGEYVRGGADAAAVADGLVRGPAVDRRAVLGESA 1670
QY 76 -----SLLSVCG 82
DB 1671 VAGVAVEGARTVFPFGQSGQWVGMRGLMGASEVFAARMRECAAVLEPHTGWDLLDVLG 1730
QY 83 ----- 82
DB 1731 EAVVDRVEVLOPASHAVAVSLAALWQAHGVVDPADVGHSGOGEIAAACVAGALSLEDAAR 1790
QY 83 ----- 82

Db 1791 VVALRSQAIAARLAGRGMA5IAPASAVETVEGVIAARNGPESTVAVGDPAVERVLA 1850
Qy 83 ----- 82
Db 1851 RYAEGRVRRRIADYASHTPHVEAIEAQLADALEGITSPTSPPWKMSTVDSGMVTEPG 1910
Qy 83 ----- 82
Db 1911 DAYWYRNLRQPVAMDTAVSELDSLFIECSAHPVLLPALDQERTVASLRTDDGGWDRFLA 1970
Qy 83 -----WSQ----- 85
Db 1971 ALAQAMTQGAADVDMWTLLIEPAPHRVLDLPTYFDHKRYWLQAPAPVTGAGIGHFFLSSAAV 2030
Qy 86 -----TI----- 87
Db 2031 LPSGDGVLTLTKLSLATHPLWLDHTVOEVVLLPGTAFLELVVRAGDEVCGDIDELVIET 2090
Qy 88 ----- 87
Db 2091 PHSPLASGAVDLTVTVDPDDTGHPRVSVHARPAGTDWTWTRHATGTLGRTEDTAPDTSNF 2150
Qy 88 ----- 87
Db 2151 POWPPTGAQPASLDEFYERLAGAGYTYGPAFOGLRAAMRAGDTVYAEVALDCAADVDRFG 2210
Qy 88 ----- 87
Db 2211 VHPALLDAAALHAGRLDAGDELELPSWTCVRLHATGAAAVRVAHSRGPAGVAVQADLDG 2270
Qy 88 -----NPEDD 92
Db 2271 RPVSVDAVLVRPSVAPSGPNLLGLEWLPVAEAHYDGADELPEGYTLITATHTDDDDDD 2330
Qy 93 ----- 92
Db 2331 PTNPHNTPRTHQTQTRVLTALQHLHTTNHTLIVHTTDPGAAVTGLTRTAONEHPR 2390
Qy 93 ----- 92
Db 2391 IHLIETHHPPLPLTQTLTLHQHLRLTNNTLHTPLHTPITTHNTTTTTPNTPPLNPN 2450
Qy 93 -----TDP----- 95
Db 2451 HAILITGSGTLGILARHLNPLTYLLSRTPPTTPRTHIPCDLTDPTQITQALHIP 2510
Qy 96 ----- 95
Db 2511 QPLTGIFHTAATHDDATLTLNLTPOHLTTLOPKADAAWHLHHHTQOPLTHFVLYSSAA 2570
Qy 96 ----- 95
Db 2571 TLGSPGOANYAANAFLDALATHRHTQGPATTIANGMHTTTLTSLQTDTEROQVRDG 2630
Qy 96 ----- 95
Db 2631 FRPLTEAEGNHLFDASLATDAPFVVAI PAEPARRRVERRTARTGEDSGRDLAVVCVAT 2690
Qy 96 -----GHAD----- 99
Db 2691 AAVLGHADASEIGAATFKDGLDLSGLRNLRLAEATGVRLSATAVFDHTPDALAA 2750
Qy 100 ----- 99
Db 2751 LABELRGDDDSQGAANAAAAAGPTVPTVAAGDEPLAIVAMACRMFGGVDTPEDLWSLVES 2810
Qy 100 ----- 99
Db 2811 GDAITEFTPTDRGWDLAALYDPDDAIGKVSVRHGGFLSGAADFAEFPGISPREALAND 2870
Qy 100 -----LVLYIT----- 105
||| ::

Db 2871 POORLVLEVSWEAFERAGILPASVRGSDAGVFMCAFTQGYGAGVLDLGGFGATGTPTSVL 2930
Qy 106 ----- 105
Db 2931 GRLSYFFGLEGPSVTVDACSSSLVALHQAAIRSLRSGECSLALVGGVTVMATTTGTFVEFS 2990
Qy 106 ----- 105
Db 2991 RORGLAPDGRAKAFADTADGTSFAEGAGVLIIVERLSDATRLGHPVLAVVRGSAVSDGAS 3050
Qy 106 ----- 105
Db 3051 NGLSAPNGPAQORVIERALDDAGLAPCDIDAVEAHGTGTRGLGPIEAQALEAAYGLDLRH 3110
Qy 106 ----- 105
Db 3111 PLLIGSLKSNFHTQAAAGVAGVIKVLAMRHGVLPRTLHVDEPSRHVDWGGDVRLLRRN 3170
Qy 106 ----- 105
Db 3171 EPWPVTGRVRRAGVSSFGISGANAHVLEAGPPAAPAPLPAADVPVEDVVPVSARTPDG 3230
Qy 106 ----- 105
Db 3231 VRDVAGRLTALTAPAAAAIGHSLATRTTAMRRAVVPARDAEAFARGSEVPVGVRTADVT 3290
Qy 106 ----- 105
Db 3291 DTRAVFVFGQSQMDGMAELLATEPVFARRLGECAELAPYTCWDLLDVIARRPGAPE 3350
Qy 106 ----- 105
Db 3351 LORVDVQOPASFAMVVAELWRAHGVAPAAVGHSGEVAACVAGVLTLDAAKVVAL 3410
Qy 106 ----- 105
Db 3411 RSLVATERAGHGMVSPADFDAAAWAGRLEVAANVGPASIVVAGAADAEBELLAATP 3470
Qy 106 ----- 105
Db 3471 HARRIADVASHAHVESIRGALLDALADLTGAPETPFSTVDEAMLDROPADAAYWYDN 3530
Qy 106 -----R 106
Db 3531 VRCVPFGAAAARLAELCHRVFVEASHPVLTALADTLAGHPNTAVTGLRRRGDGGARR 3590
Qy 107 F----- 107
Db 3591 FTRSLAELWVRGVPVSNWPFGBELRGVPLPTYPFRDRYWDAPAGTSGHPLGLSLVERAD 3650
Qy 108 ----- 107
Db 3651 GEGAFATLFSRRQRPWLADHEVDGRIIVFGSALVELLAEGARLGTPTIAELTIVAPVV 3710
Qy 108 ----- 107
Db 3711 VDEERDEIQLTVATEVSGRRVRLHSRTGAGPWTGATGALSADTRDPEPVEWPADAE 3770
Qy 108 ----- 107
Db 3771 PVDLTGFDYDPLTISYGPAPFRAMTAMWAGOGRAVASVRLGEQLTDARYGLHPVLDAVHA 3830
Qy 108 ----- 107
Db 3831 LGTLFADPERRRLAFSWSGVRIHARAATTQORVULLERADGDTIRILATDSHGAPVLDVDGL 3890
Qy 108 ----- 107
Db 3891 TVRAAEFGTEALFEVAVWSVPASVPBGWTHLADVPEGDVLEGAGEGVSDPPPPVVVL 3950
Qy 108 ----- 107
Db 3951 AVEFGPDDSSPGVRARELGRDLLATVTRTLAAAPRAESRIVVTRTGDPAQOALGGLVRT 4010

```
Oy 108 ----- 107
Db 4011 AGTEHPRVGRIBADEITPATVEAGLAGDESHVRVTDGTVRAARLRRVTASTGASPLSG 4070
Oy 108 ----- 107
Db 4071 GTVLVTGGTGLRLLVDHLLTHAEAEVVVASRGRPGCAPADDRVRYVAADVTRDEL 4130
Oy 108 ----- 111
Db 4131 AALVDGVAERLCVVMHAGIVDDAVVATMRPQOWDAVLRKADVAWQLHQLHTRDLELAAF 4190
Oy 112 ----- 111
Db 4191 VLYSSISATFGGAGQANYATGNAFLDALARHRRDOGLPAVSLAWGLWDEADGMGRLAAT 4250
Oy 112 ----- 111
Db 4251 DLTRIARGGMI PMTAEQGLELFDAAALHSDRAALVPVRDLAAVAASDQVPPILRDLVPNA 4310
Oy 112 ----- 111
Db 4311 RRAAAPTAAARDMLVRTSAALVLGHRDAHAVEPTRAPKEVGFDSILTGVELRNRLADATG 4370
Oy 112 ----- 111
Db 4371 LTLPATLVFDHTAQALAAHLDELAGARTATRRRTAPARRHDEPLAIVGMACRLPGGVA 4430
Oy 112 ----- 111
Db 4431 SPDDLWRLLESGGDGITAFPADRGHDVDALYDPDEHSGTSTVRHGGFLTGAGDFDAGFF 4490
Oy 112 ----- 111
Db 4491 GISPREALAMPQORVLETSWEALEHAGIDPHTLRGTTGTGVFMGGYFYGYSGADRGDF 4550
Oy 112 ----- 111
Db 4551 GATSTQTSVLSRLSYFYCLEGPVTVDTACSSSLVALHQQSURTGECSLALVGGVTV 4610
Oy 112 ----- 115
Db 4611 MASPSGFVDFSQORGLAPDGRCKAFAEADGTAFAGSGVLVVERLSDAERHGRVLAUV 4670
Oy 116 ----- 115
Db 4671 RGSVYNQDGASNGLSAPNGPSOERVIRQALANAGLQPSVDIAEAHGTGTRLGDPLEATA 4730
Oy 116 ----- 115
Db 4731 LLATYQDORATPLLGLSLKSNIGHTQAAAGVAGIIMVLMRHHGTLPTLHVDTSPSHVD 4790
Oy 116 ----- 115
Db 4791 WSAGTVELLTDARTWPETGRPHRAAVSFGVSGTNAHVIIESHPRPTPAPDTGSGSTHVPV 4850
Oy 116 ----- 115
Db 4851 LLISARTPEALDEHTTRVRAFLDAEGEDEERAVASALLTRTAFTHRAALIGTLITGTAEP 4910
Oy 116 ----- 115
Db 4911 DRRLVWLFSGQSQRPQMGDELAAYDVFARTRRDVLALQVPAGLDVHDTGYAQPAVFA 4970
Oy 116 ----- 115
Db 4971 LQVALSAQLDANGVRPDLVGHSGIGELAAAYVAGVMSLDDATELVASARLMQALPPGGA 5030
Oy 116 ----- 115
Db 5031 MAASASERDALPLLCEGVETAAVNGPASIVLSGDEDAVLDAVRAARLGRFTRLRTRSHAFHS 5090
```

```
Oy 116 -----ROVR----- 119
Db 5091 ARMEPMLDEFRDVAERLTYHEPKLPMAAGADCATPEYVWVRQVRDTRVFAEOVAAVDGAAL 5150
Oy 120 ----- 119
Db 5151 LEIGPDRNLARLVLDGIPVLHGEDEARSAMTALARLHTGGVAVDWPVEVIGAAPTDLPHLPT 5210
Oy 120 ----- 119
Db 5211 YPFERTRYWLSRAAGDAAAPAGOLPVAHPVLTAAMVMPGTGDLVLTGCRVEATDPLATRV 5270
Oy 120 ----- 119
Db 5271 HPGAARAVLLDLRSRRATKRAAAPWRRSPMTPRSALPQSGGLALSVTVAAPGEDGRRAVA 5330
Oy 120 ----- 119
Db 5331 VHARPGTGENTEHATGILARVARVASAVPETSPWPATARPFDVGLADRLARAGHTDG 5390
Oy 120 ----- 119
Db 5391 PALPRPRAAWADDDAVHAEVALADEQHAERYGLHPALLGAALALAGECADLPSAFDDV 5450
Oy 120 ----- 119
Db 5451 RVHATGATTVRVAVTATGIHLADETGSVATVGAVRRRPLTVEGAVPGLLRPLDAEIAEL 5510
Oy 120 ----- 119
Db 5511 PPTATTGGLLDDPVVPDVILPAHGTGGGPGLETGRLGARVLAALRSFLTDDRIVADAVLA 5570
Oy 120 ----- 119
Db 5571 VHTGPGLPAAAGLVRTAQAEHPRIVVVDABPDTAAPLLAAAAAGLGEPOQVVLREGRA 5630
Oy 120 ----- 119
Db 5631 ARRLTPAVPSGDAPELDPDGTVLTITGSGTGLAGIVARHLVGHYGVRRLLMLSGGTASDV 5690
Oy 120 ----- 119
Db 5691 PGAEVTAACDIADROELASALAGIDPAHPLTAVVHTAAVDDGVLTALTPERLDTVLRP 5750
Oy 120 ----- 119
Db 5751 KADGAWHLHELTKDQDLAALCSTPRPECSATRDRTTPRRTLDELAEORRAGGLPALS 5810
Oy 120 ----- 124
Db 5811 VAMGLWEPESGLTAGAGDRMRDGTALTSTERGLALLDAALRSPEAALVAADPGGLGSSA 5870
Oy 125 ----- 124
Db 5871 LLRAPHGPRRAAGASALADRIAGLSAADAEBKAAALAVVRECAAAVLGHEGPARIATAT 5930
Oy 125 ----- 124
Db 5931 FKELGVDSLTAVRLRNAFTATDVLPAATAVDFPTPQAAKLTAALTGCRVPARARATK 5990
Oy 125 ----- 130
Db 5991 AVCGHDEPLAIVGMACRLPGGVASPDLLWRLLESQGGGITAPPADRGWDVDALYDPPDH 6050
Oy 131 ----- 135
Db 6051 SGTSTVRHGGFLTGAGDFDAGFFGISPREALAMPQORVVULETSWEALEQAGIVPGTLRG 6110
Oy 136 ----- 140
Db 6111 SDTGVFMGAFSDGYGLGTLGGFGATGTQTSVLSGRLSFYFYLEGPVAVTDTACSSSLVA 6170
Oy 141 ----- 140
```

Db 6171 LHQAGQSLRTGCSLALVGGVTVMASPGGFVFSQQRGLAPDGRCKAFADGTAFAEG 6230
QY 141 ----- 140
Db 6231 SGVLVVERLSAERHGRHLAVVRGSAVNQDASGLSAPNPSQERVIROALANAGLQP 6290
QY 141 ----- 140
Db 6291 SDVDAIEAHGTGRLGDIETATALLATYGOHRTPTLLGLSLKSNIGHTQAAAGVAGIIM 6350
QY 141 ----- 140
Db 6351 VLAMRHGTLPTLHVDTFSSHVDWTTGCVELLTOPRDPWPQTRPHRAGISFGVSGTNAHV 6410
QY 141 ----- 140
Db 6411 IIEAHVAEPPALNSPGDPSVAEPGASLFATERPLPVSARTSEAVDGOVQBLREHLAA 6470
QY 141 ----- 140
Db 6471 HPRDDPRVAAALLATRTDFPHRAVLGDDVVVTGTALTAPRTVFVPGQSQWLGMGREL 6530
QY 141 ----- 140
Db 6531 MAESPVFARMQCADALAEHTGRDLIAMLDDPAVKSRVDVWHPVCWVMVSLAAVWEEA 6590
QY 141 ----- 140
Db 6591 GVRPDVAVHSGEIAAACVAGAISLEDGARLVALRSALLVRELARGAMGSIATAAAR 6650
QY 141 ----- 140
Db 6651 IDGVWVAGRNGTATTIVSGRDPDAVETLIADYEARGVWVTLVLVDCPTHFTFVDPVDELO 6710
QY 141 ----- 142
Db 6711 RIVAATTSRAPEIPWFSTADERMIDAPLDDEYWFNRNRPVGFAAVAAREPGDTVFIE 6770
QY 143 ----- 142
Db 6771 VSAHPVLLPAINGTTVGTLRGGAGRVLDLSAKAHTVGVAWDMVTVAATGAADDAASV 6830
QY 143 ----- 142
Db 6831 TAHDTGTAHDLPTYAFHRYWIEPATGTASGLGLDVAHDHPLLAASVALPDSDEVLLTG 6890
QY 143 ----- 142
Db 6891 RLSLATHPMLAGHTVDSGEVLLPGPVFVELAGRGADAEARCDLDELVTETPLALPATGAVQ 6950
QY 143 ----- 142
Db 6951 VRVTAAADAGRAVRARIHARHGGRTWTRHASGIAAIATTESATADGPPPARAERIDI 7010
QY 143 ----- 142
Db 7011 ADFYRLDEARTSSAPSSGMSAAWSHGDTLPAEVALDGTLDARDAARYTLHPALLVTALQA 7070
QY 143 ----- 142
Db 7071 GSLAAGSDAGVRLPFAFTGVRVHSSGATKARVTFAAGDGGARVHIADELGPVVAEIGSF 7130
QY 143 ----- 142
Db 7131 VTRPLAEADRGDVRLYRRYRTWTVRAPAAPGTTAVRYADLGDGATSDATSPDAASPDVVL 7190
QY 143 ----- 142
Db 7191 MRASDPAEVRAALDDPRTDATLVVSAEAGSVAGAVAAALLDAEPGRLLVLTQTDVTPR 7250
QY 143 ----- 142

Db 7251 RAAALARLDEPHRLTGTORLEAPRLVPAPTPTTTAAAPYGDVLLDGGSEBELARRLGEHGAE 7310
QY 143 ----- 142
Db 7311 VIHHKPGKLPETPVSAVVHGAADSPWELHRLTRDLPLTAFLVLLVPPGGAAPWEALAE 7370
QY 143 ----- 142
Db 7371 LRRTEGLPAVAFTATEDROTDLDDVACATGEAVVATAPRRRGRDPSPLWRPVQRPARRSA 7430
QY 143 ----- 146
Db 7431 SDGSLRLERLPDLSPKDQEQVLLDVVRDTAATLLGHADARAVTATTAFKDLGV 7483
RESULT 12
A:5086
HC-toxin synthetase - fungus (Cochliobolus carbonum)
C:Species: Cochliobolus carbonum
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 03-Nov-2000
C:Accession: A45086
R:Scott-Craig, J.S.; Panaccione, D.G.; Pocard, J.A.; Walton, J.D.
J. Biol. Chem. 267, 26044-26049, 1992
A:Title: The cyclic peptide synthetase catalyzing HC-toxin production in the filament
A:Reference number: A45086; MUID:93100128; PMID:1281482
A:Accession: A45086
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-5232 <SCO>
A:Experimental source: SB111
A:Note: sequence extracted from NCBI backbone (NCBI:120884)
C:Superfamily: Cochliobolus carbonum HC-toxin synthetase; acetate-CoA ligase homolog
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:277-743/Domain: acetate-CoA ligase homolog <ACLI>
F:784-854/Domain: acyl carrier protein homolog <ACP1>
F:1882-2366/Domain: acetate-CoA ligase homolog <ACP2>
F:2397-2464/Domain: acyl carrier protein homolog <ACP2>
F:3036-3526/Domain: acetate-CoA ligase homolog <ACP3>
F:3549-3618/Domain: acyl carrier protein homolog <ACP3>
F:4186-4655/Domain: acetate-CoA ligase homolog <ACP4>
F:4680-4750/Domain: acyl carrier protein homolog <ACP4>
F:818,3583,4715/Binding site: phosphopantetheine (Ser) (covalent) #status predicted
Query Match 28.7%; Score 222; DB 2; Length 5232;
Best Local Similarity 1.8%; Pred. NO. 59;
Matches 87; Conservative 20; Mismatches 36; Indels 4702; Gaps 27;
QY 2 AGCI----- 5
Db 304 AGGVCVPIDPRYPVERIRDIIRTNATIALVGAGKTAALFKSADTAVOTINATKDI PHGL 363
QY 6 ----- 5
Db 364 SDTVOSNTKIDDPAPGLFTSGTGVKCIIVTHSOICTAVQAKDRFGVTSRVLQPS 423
QY 6 ----- 5
Db 424 SYTFDISIADTFALFYGCTLCIPSEDRMSNLQDYWVSRPNWAVLTPTVSRFLDPGVV 483
QY 6 ----- 5
Db 484 KOFISTLI FTGEASREADTPWIEAGVNLVNYVGAENTLITTRIRKGSNIYG VNV 543
QY 6 ----- 5
Db 544 TRTWTVDSGACLVPGVSGIGELLIESGHLADKYLNRPRDTEAFLDPLWPNYEGDSVR 603
QY 6 ----- 5
Db 604 RGRFRYRTGDLVRYCDGSLICVGRSDTQIKLAGORVELGDVEAHLQSDPTTSQAAVVPF 663
QY 6 ----- 5

Db 664 RSGPLEARLIALVTGNKDGTFHNOQSLPKPAFAQCPDVLVKYATSSLOQRPSYMWPSV 723
 Qy 6 ----- 5
 Db 724 WLGIDFLPMSVSGKLDRVAVLQDQLESLSPSDYAEILGTTGLEVDPGGAASSVASDSLDR 783
 Qy 6 ----- 5
 Db 784 MNDDSLLLTACSRVNLNLPAGKISYSQSFHAGGDSITAMQVSSWMKRTGKRIGVKOLLV 843
 Qy 6 ----- 5
 Db 844 SPSISTAASCIKSAQDGRNFVAVRPGORI PVSPQIKLFFQTAEASKSWNHYSFLPRI 903
 Qy 6 ----- 5
 Db 904 DOPIKQTIEDAISLVMPORHPMLQARFERTBEGDWYQIPIIDVERRASVEVIGSLSTDDR 963
 Qy 6 ----- 5
 Db 964 EAMLRARQSIDLTEGLRQCFNNNVDEASRLFFVVIHHAUVDLVSWRIIMEEHL 1023
 Qy 6 ----- 5
 Db 1024 ATDSTPDRGEAYQESVPFLAWCQVQAEAVKOIPVDRTPVLPKPTADFGYWGKLDENV 1083
 Qy 6 ----- 7
 Db 1084 YGNTVERKIPLGHSITEDLLYKCHDSLTKTIDVLLAAVLVSFRFSFLDRPVPAVFNESH 1143
 Qy 8 ----- 7
 Db 1144 GREPGEDAVDLRTVGWFTTISPVVPEVSPGDILDVVRVKDYRMATPNNGDFYFSTK 1203
 Qy 8 ----- L 8
 Db 1204 YLTQSGIKLFEDHLPAILFNTEGRYQAMESEQTVLKPESWHAGEASKDQDPLRRFLCF 1263
 Qy 9 ELLVAVGPD- 17
 Db 1264 EISTAVLPDQGLHLTCSMKNRHOGRIRLWLDTLPPAIGEIVSSALASQPLTSDVE 1323
 Qy 18 ----- 17
 Db 1324 LLRLDYSSLDILKKSILSIPAVQTLDDLEGVYVPGPMODALFLSQSKSQDGAYEVDFTW 1383
 Qy 18 ----- VFOAHOED- 25
 Db 1384 RVATSLONSOPAVDIGCLVEAKDVTVALHAALRTVILESSLPATGILHGVLRSHDPDIV 1443
 Qy 26 ----- TE- 27
 Db 1444 ILDRVDTAAITLDSYPPPTBEGIALIKRPPHRLICTTIEGSLVLIKQVNLHVFDMGS 1503
 Qy 28 ----- RYV- 30
 Db 1504 TDKIIQDLKAVTCRHSNKLPHSESKLHDGTGNRTPKPLAEFIRYIRDPQRKQDSIN 1563
 Qy 31 ----- LTNLNG- 37
 Db 1564 YKNAURGATTCSPFLFDIOITSEKAMPQSWASVPIPLCVDSKELSKTLANLIGTMTM 1623
 Qy 38 ----- 37
 Db 1624 FQVWAILVLIYSONQSVFGYLTSGRDAPVDGIDSAVGNFIAMLVCFDDDDGVHTVA 1683
 Qy 38 ----- 37
 Db 1684 DMARKIHNASANSISHQACSLABIQDALGLSTSTPLNTAFTYLPKPTNVKAGEPEHHL 1743
 Qy 38 ----- AELLR 42
 Db 1744 CFEELSMDSPTFEDTLFVPTQESNEVSAHLDFKLSYISQAYATSIATVAHILSELVH 1803

Qy 43 DP----- 44
 Db 1804 DPYRALNTLPIVSEHDTAIIIRSNWDLFPATECIHETSRKVVEHPQREAIACSWGSLT 1863
 Qy 45 ----- SILGAQFRVHLVKM- 57
 Db 1864 YAEISLDSORLSLHLVSLGIVGTKIPICFEKSMWITVILAVVQAGGVFVLEPGHPES 1923
 Qy 58 ----- VILTEPE- 64
 Db 1924 RLSGIIKQVQAEILLCSPATSRMGALQNISTOMGTBFKIVLEPEFIRSLPLPKPNHQP 1983
 Qy 65 ----- GAP- 67
 Db 1984 MVGLNDDLYVFTSGTGVPKGAVATHQAYATGIYEHAVACGWTSLGAPRSLQFASYSF 2043
 Qy 68 ----- NITANL- 73
 Db 2044 DASIGDIFTLAVGGCLCIPREEDRNAGITTFINRYGVTWAGITPSLALHLDPAVPTL 2103
 Qy 74 ----- 73
 Db 2104 KALCVAGEPLSMVVTWMSKRLNLINMYGPTVACIANOVCTCTTTTSDICRGYRATT 2163
 Qy 74 ----- 73
 Db 2164 WVQPDNHNLSLPIGAVGELIIEGSLCRGYLNDPERTAEVFIIRSPSWLHDLRPNSTLYK 2223
 Qy 74 ----- 73
 Db 2224 TGDLYRSADGKIIFIGRKTQVKNMGORFELGEVEHALQLOLDPDGPPIVOLLKRTQS 2283
 Qy 74 ----- 73
 Db 2284 GEPDLLIAFLVGRANTGTGNSDEIFIATSTSSLSEFTVIKKLODQARAMEVLPLMPVP 2343
 Qy 74 ----- TSSLLS- 79
 Db 2344 QAYPIEGGIPLTAAGIKDRMLRKLCEPNRNDLISFTSKALSTSVKDAETTTVDRL 2403
 Qy 80 ----- 79
 Db 2404 ARIWEKVLGVKVGRESDFSSGSGNSMAIALRAEAQRSFTLFLVADIPTNPLRLADMAKL 2463
 Qy 80 ----- 79
 Db 2464 FSHQSVSPSSSTLRTKVPISLQKSSGLQTAAAPVNSGSPVRRCKENIIDCPVAFEYE 2523
 Qy 80 ----- VCG- 82
 Db 2524 EGSDTOLKEASRICGIGSSRSIEDVFPCTPMQEALVALSLIPGAQASYALHAAPLRLPGL 2583
 Qy 83 ----- WSQ- 82
 Db 2584 DRNFRSAWESTVKAQPIILRSRIISGSGNSVVVTSATDSIPQLDVSGLDTFLEQQLQVG 2643
 Qy 83 ----- 85
 Db 2644 PAPGAPLRLAFVYSKADDCDYFVISAHAIYDGMNLNIMSQVLAALYTNLNGELPPGSPF 2703
 Qy 86 ----- 85
 Db 2704 KHFAENLNLVOSKLDSEDFWRKLLVKPDOSFRFPDVPVGHKPATRCTTTFHFPFSMOSK 2763
 Qy 86 ----- TIN- 88
 Db 2764 IGTANTCINAAWAITLAQYSSNKTVNFVTLWGRDFPMIDIEHMTGPTITVTPROVNV 2823
 Qy 89 ----- 88
 Db 2824 PESSVAEFLDQLQKSLAVVLPQHGLHRIQALGPBARQACDFSTLLVNNHGSSISWSSEL 2883

QY 89 ----- 88
Db 2884 EAADIVPPLRSSLDYAYPMVVEVENASSDTLDIRVHSDPDCIEVOLLERLMEQFHNQ 2843
QY 89 ----- 88
Db 2944 TLCRAASFDPKRIAEMLDDTATTHLRTLFSWNSRVKSDPDVAAIAVHKLEETAQSQA 3003
QY 89 ----- 88
Db 3004 ESAIVAHGOLSYMOMDRCADVLARQIRKTNMISAQSPFVCIHLLRSATAVWSMLVKA 3063
QY 89 ----- 91
Db 3064 GGAFMPVDISOPRSRLQNLIEESQAKLVLLPESANALATLSGLTKVIPVLSLVOQIT 3123
QY 92 ----- 95
Db 3124 DNTTKDEYCKSGDTPSSPAYLLYTSGTSGKPKGVVMEHRAWLSGFTCHAEYMGFNSCT 3183
QY 96 ----- 95
Db 3184 RILOFSLMFDLSILEIAVAVYAGCLFIPSDKERVNNQDFTTRINDINTVFLTPSIGKL 3243
QY 96 ----- 95
Db 3244 LNPXDLNPI SFAGFIBPMTRSLIDAWTLPGRRLLVNSYGPTEACVLVTAREISPTAPHDK 3303
QY 96 ----- 98
Db 3304 PSSNIGHALGANIIVVEPORTALVPIGAVGELCIEAPSLARCYLANPERTEYSFPSTVLD 3363
QY 99 ----- 105
Db 3364 NWQTKGTRVYRTGDLVRYASDGTDLGLRGDGOIKLGRGRIELGEIEHHIRLMSDDPR 3423
QY 106 ----- 105
Db 3424 FHEASVOLYNPATDPRDATVDVQMRPEPYLAGLLVLDVLSLMSRSDSHVNLINANTSEN 3483
QY 106 ----- 105
Db 3484 QTLVTELKSLRGVLPHYVMPLEHFVAVSRULPTGSSGKLDHAFVACLRELTAPLDGNFPK 3543
QY 106 ----- 105
Db 3544 VEQVLTINESVLROWGTVLAMPHSIQRGDDFFSLGSSISAMRLVGLARSSGHKLOHE 3603
QY 106 ----- 109
Db 3604 DIFMCPRLADWAGQISFVQEASVSPPTSPTIKFDLLDDCEVDEVIDHILPOLDMNKELIE 3663
QY 110 ----- 109
Db 3664 DIVYPTPLQESLMAATARHGEAYTMIQSIITVLAQAQLKKAMDVVRFDPEVLRTRIALG 3723
QY 110 ----- 109
Db 3724 PSQOALQVVVKHELSWESFPFISQKDFYRSLGKPLARLAVITQALDTKQPI SHGT 3783
QY 110 ----- 109
Db 3784 REARTKNSQOTVMVVVGAHHSIYDAHVLMSIWRRLYREFIGSQADGILEAETSRSEGVVP 3843
QY 110 ----- 113
Db 3844 FKSYYEKLRLCKNDSELLFWKEKLRGVSSQFPFASPVRVLEHQPSATQTLITKVSLEPT 3903
QY 114 GNRQVRGVT----- 122
Db 3904 SSRKKGATVATVAYAAWALTIHYTADPDVVGATLSGRETMAGSISHPESIAIPTIT 3963
QY 123 ----- 122

Db 3964 VPLRIIIDFQTVVSDFLSTLQKDIVRAAYFGQMMGLNSIAHIDNDCRDCAGCFKSIIVQV 4023
QY 123 ----- 122
Db 4024 PDEGENHVGRAANPFPOMLESIGHPPAPLVVEVESESTDVILIRMAVDVPLVPEKLAHFI 4083
QY 123 ----- 122
Db 4084 SDTFTTMSNLNANPKAKVESIPALSEAHLAELDVTCPEWILGKAKDEKIRTESHQCLQ 4143
QY 123 ----- 125
Db 4144 DLVCRRAQSPNSQAIDSDWGSISYHELDGLSILAEHLSQLGVRPEAPVCLLFEKSKWA 4203
QY 126 ----- 125
Db 4204 VVAMIGIIRKAGCFVPLDPSYPHERLEHIISETGSSVITVSAAYSKLCLSLSVRGIVCDG 4263
QY 126 ----- 125
Db 4264 SVFSSTKKPLPSTADSPSPFSVVRPNQAAVILFTSGSTGKPKGVVMEHHSVCSALIALGKR 4323
QY 126 ----- 128
Db 4324 MGLGPOSRLQFNYSYWFVWMLLDIFGTLVYGCLCIPKEEORMNSLGSWQKFKVNTMLL 4383
QY 129 ----- 128
Db 4384 STSVRLMOPADTPSLTCLTGEAVLQSDVDRWAPKLHLIAGYGTETCIMSUSGELTP 4443
QY 129 ----- 128
Db 4444 SSPANLIGKPVSCQAVVINPLKETELAPYGATGELYIQGTVARGYLHDDVLTSKAFIVD 4503
QY 129 ----- 128
Db 4504 POWLTGYKTENQWSRRAYKTGDLVFWPQSNLYVYRRKSSQVKIRGORVELAEIEVI 4563
QY 129 ----- 128
Db 4564 RQHIPPDTVCVLLSSDDQNTRIILGAVLIGDRLAGPDELEVIGYMDLKSIIIPAL 4623
QY 129 ----- 131
Db 4624 EASLPHHMIPEAYVPFVQLPTLGLSGKLDKTVRRVAGPLAFSLPOASARHPNQPTVHTQ 4683
QY 132 ----- 131
Db 4684 KLLRLQWCKILPOLDESAVNKQDNFLGIGGDSIAAIKLVALLRQHGISLAWAEIFTRPTL 4743
QY 132 ----- 131
Db 4744 EAMSSLIDEHNFVSHAGILSDVTRNTSGVMROTTNLIAGRHSMAVEKSRCDNSTLPCT 4803
QY 132 ----- 131
Db 4804 EYQOMFLAGTEAFTGAHSAQFIERLPEKIDLDRLQAADFHCADWYPNLRTQIHKDADTGR 4863
QY 132 ----- 134
Db 4864 LLHDISPIGVKVPSWCHYSDDLNTVLSHDKKPPPLGDLPLHVRTIMRHRDPTESMLVMTL 4923
QY 135 ----- 140
Db 4924 NHAAYDAWSLRMLMEHTEAYANPDYEPSYSLGWTAFVLHTENTKEASRSFWSYLSVDK 4983
QY 141 ----- 140
Db 4984 PARLMFNVLVSNPRODRLYEARINIPKRVLSQATATVLLAGTLVARVCDTRDVILA 5043
QY 141 ----- 140

Db 5044 HLTGRTLPLAGIENCPOPTITKVPLRIPLMDQDLVTLEDSVAKKITAEMLRVMPEHS 5103
Qy 141 -----GFDL 144
Db 5104 GLSAIREFIPOAEGTTTSSGKFHAGSVLGRPLDLIVHPKGGDL 5148
RESULT 13
A88852
protein unc-22 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: A88852
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: A88852
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-6831 <STO>
A:Cross-references: GB:chr_IV; PIDN:CAA98081.1; PID:g3881830; GSPDB:GN00022
C:Genetics:
A:Gene: unc-22
A:Map position: 4
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;
Query Match 28.6%; Score 221; DB 2; Length 6831;
Best Local Similarity 1.3%; Pred.No.1.7e+02;
Matches 81; Conservative 28; Mismatches 32; Indels 6181; Gaps 25;
Qy 3 GGILHL----- 8
Db 146 GGLYHAIFSDLGDTYLCOLEIRGPSSDAGQVRCNIRNDQGETNANLALNFEEDPSER 205
Qy 9 ----- 8
Db 206 QBRKRSTASPRSPGSRSPSPKMSKMSREGTPKRTLKPREGSPSKLRSRTSTPVNE 265
Qy 9 ----- 8
Db 266 EVSQSESRSRSTDKMEVDQVSGAKRKPDPGLPPGGDEKLRAGSPSTRKSPRSKASAP 325
Qy 9 ----- 8
Db 326 TSPRKSSAGGAASGTTGASATSATSATSGSSASSDASRDKYTRPPIVLEASRSQTGRIGG 385
Qy 9 ----- 8
Db 386 SVVLEVQWOCHSSTIIWYRDGTLVRNSSEYSQSFNGSIKLVNKLTEKSGLYKCHAK 445
Qy 9 ----- 8
Db 446 CDYGEQSSAMVKIESDVVEELMKHAKDAEYQKEBOKSQTLQAEKRVARRSKSKS 505
Qy 9 ----- ELL 11
Db 506 KSPAPQAKSTTSQGRQEAQVEHKKRSSVRDPDEESQLDEIPSSGLTIPERRRELL 565
Qy 12 VAVG----- 15
Db 566 GQVGESDDEVSISILSPAGGPKRRTDDPKPKVSIAPVSTNKSDDPSTPRRSI 625
Qy 16 ----- 15
Db 626 DMRRESVQIELEKTSTPLVPSGASGAPKIVEPVNTVVENETAILTCYVSGSPAPTR 685
Qy 16 ----- 15
Db 686 WFKSREVISGGRFKHTDCKEHTVALALLKCRSQDEGPYTLTIENVHGTDSADVLLVT 745

Qy 16 ----- 15
Db 746 SDNGLDFRAMLKHRESQAGFQKDEGGGAGGEGGKPMTEAERROSLFPGKKVEKDIP 805
Qy 16 -----PD 17
Db 806 LPEKTVOQVDKI CEWKCTYSRPNKIRWYKDRKEIFSGGLKYKIVIEKNVCTLIINPE 865
Qy 18 VFOAHQEDTERY-----VLT----- 32
Db 866 V-----DDTKYTCEANGVTHAQLTVLEBPMKYSLNPLPNTQEIYRTKQAVLTCKVNT 920
Qy 33 ----- 32
Db 921 PRAPLVWYRGSKAIOEGDPRFIIIEKDAVGRCTLTIKEVEEDQAEWTARITQDVFSKVQV 980
Qy 33 ----- 32
Db 981 VYEPHRTFVVPKMSOKVNESDLATLETVDNDKDAEYVVMHDKRIDIDGVKFKVSSNR 1040
Qy 33 ----- 32
Db 1041 KRLIINGARIEDHGEYKCTTKDDRTMAQLIVDAKNKFI VALKDEVIEXDDVTLMCOTK 1100
Qy 33 -----NLNIG----- 37
Db 1101 DTKTPGIWFRNGKOISSMPGGKPFETOSRNGTHTLKICKIEMNEADYVEIDQAGLRGCVN 1160
Qy 38 -----AELLR----- 42
Db 1161 TVLEAEKRPILNWKPKKIEAKAGEPCVVKVPFOIKGTRRGDPKAOILKNGKPIDEEMRKL 1220
Qy 43 ----- 42
Db 1221 VEVIIKDDVAEIVFNQOLADTGKWALELNSAGTALAPFELFVKDKPKPKPLETKNV 1280
Qy 43 -----DPSLGA----- 48
Db 1281 TAEGLDLVWGTPDPDEGAPVKAYIIEMQEGRSGNWKVGETKGTDFVKVLDLKEHGEYKFR 1340
Qy 49 ----- 48
Db 1341 VKALNECGLSDLTGESVLAKNPYGVPGKPKNMDAIDVDKDHCTLAWEPPEDGGAPITG 1400
Qy 49 ----- 48
Db 1401 YIIRERKSEKDWHQVGTQKPDCCELTDKKVVEDKEYLYRVKAVNKAGPGDPCDHGKPIK 1460
Qy 49 ----- 48
Db 1461 MKAKKASPEFTGGGIKDLRLKVGETIKYDVPISGEPLPECLWVNGKPLKAVGRVKNSE 1520
Qy 49 ----- 48
Db 1521 RGHIMKIENAVRADSGKFTITLKNSSGCDSTATVTVVGRPTPPKGPLDIADVCADGAT 1580
Qy 49 ----- 48
Db 1581 LSNPPDDDDGDLTGVIYEAQMDNKGKYEIVGKVDPNPTTLKYNGLRNKNGYKFRVKA 1640
Qy 49 ----- 48
Db 1641 VNNEGEPLSADQYTOIKDPWDEPGKPRPEITFDADRIDIAWEPHKGCGAPIEEYI 1700
Qy 49 ----- 48
Db 1701 VEVRPDPTKEMKEVKRVPDPTNASISGLKEGKEYQFRVAVNKAGPGQSPSEKQLAKPK 1760
Qy 49 ----- 48
Db 1761 FTPAWLKHDNLKSIITVKACATVRWEVKTGGEP1PEVKWPKGNQOLENGIQLTIDTRKNEH 1820
Qy 49 ----- 48

```
Db 1821 TILCIPAMRSVGEYRLTVKNSHGAEKANTVLDPRSPKNGPLEVSDVFEDNLNSW 1880
QY 49 ----- 48
Db 1881 KPDDGGEPIEYVEVEKLDATGRWVPCAKVKDTKAHIDGLKKGQYQFRVAVNKEGA 1940
QY 49 ----- 48
Db 1941 SDALSTDKDKAKNPYDEPGKTGTPVDWDADRVSLEWEPKSDGGAPITQYVIEKKGK 2000
QY 49 -----OFRV----- 52
Db 2001 HGRDQECGKVGSDQTNAEILGLKEGEYQFRVAVNKAAGPEASDPSRKVAKPRNLKP 2060
QY 53 ----- 52
Db 2061 WIDREAMKTTIKVGNDFEVPVRGEPKKEWIFNEKPVDDOKIRIESDYKTRFVLR 2120
QY 53 -----HLVKMVL----- 60
Db 2121 GATRKHAGLYLTATNAGSGDKHSEVIVLCKPSSPLGLEVSNVYEDRADLEWKVPEDD 2180
QY 61 ----- 60
Db 2181 GGAPIDHYIEKMDLATGRWVPCGRSETTKTTPVNLQPGHEYKFRVAVNKEGESDPLTT 2240
QY 61 ----- 60
Db 2241 NTAILAKNPYEPVGKDKPELVDWDKOHVLDLWNPADGGAPIEAFVIEKDKNGRWEA 2300
QY 61 ----- 60
Db 2301 LVVPGDOKTATVPLNKEGEYQFRISARNKAGTGDPDPSDRVAVKPRNLAPRIHREDLS 2360
QY 61 ----- 60
Db 2361 DTTVKVATLKFIVHIDGEPAPDVTWFSNGKIGESKAQIENEPEYISRFALPKALRKQSG 2420
QY 61 -----TPEG----- 65
Db 2421 KYTITATNINGTDSVTINIKVKSXPTKPGPIEYTDVDFEDRATLDMKPPEDDGGEPYEFY 2480
QY 66 ----- 65
Db 2481 EIEKXNTKDGIIWPCGRSGDTHFTVDSLNGDHYKFRVAVNSEGSPDPLETETDILAKN 2540
QY 66 ----- 65
Db 2541 PFDRPRGRPEPTDWDSDHVDLKWDPPLSDGGAPIEBEYQIEKRTKYGRWEPAITVPGGQ 2600
QY 66 ----- 65
Db 2601 TTATVPOLTPNEEYFRVAVNKGSPDPSDASKAVIAKPRNLKPHIDRDLAKNLTKAG 2660
QY 66 ----- 65
Db 2661 QSIISFDPVSGEPAPTVTHWHPDNRNGRGRVLDNPEYOSKLVVKQMERGDSGTETIK 2720
QY 66 ----- 65
Db 2721 AVNANGEDATVKINVIDKPTS PNGPLDVSVDVHGDHVTLNWRAPDDGGIPIENYVIEKY 2780
QY 66 ----- 65
Db 2781 DTASGRWVPAAKVAGDKTTAVVDGLIPGHEYKFRVAVNAEGESDPLETFTGTLAKOPFD 2840
QY 66 -----AGNIT----- 70
Db 2841 KPGKTNAPETIDWDKXHDVLEWKPPANDGGAPIEBEYVVMKDEPSPFWNDVAHVPAQOTN 2900
QY 71 ----- 70
```

```
Db 2901 ATVGNLKEGSKYEFIRAKNKAGLGDPPSDASAVAKARNVPPVIDRNSIOEIKVKAGQDF 2960
QY 71 ----- 70
Db 2961 SLNIPVSGEPTPTIITWTFTFESTPVESDDRMKLNNEGDKTKFHVKRALRSDTGTVIKAENE 3020
QY 71 ----- 70
Db 3021 NGDTAEVKVTVLDHPSPSPRGLDVTNIVKDGCDLAWKEPEDDGGAEISHYVIEKODAAAT 3080
QY 71 ----- 70
Db 3081 GRWTACGESKOTNFHVDDLTQGHYKFRVAVNRHGDSDPLEAREAIIAKDPFDRADKPG 3140
QY 71 ----- 70
Db 3141 TPEIVDMKDHADLKWTPPADDDGAPIEGVLVEMRTPSGDWVPAVTVGAGELTATVDGLK 3200
QY 71 ----- 70
Db 3201 PGQYQFRVAKLNKAGESTPSDPSRTMVAKPRHLAPKINRDMFVAQVRKAGOTLNFVNV 3260
QY 71 ----- 70
Db 3261 EGEPAKIEWFLNGSPSSGGNTHIDNNTDNNTKLTSTARADSGKYKIVATNESGKDE 3320
QY 71 ----- 70
Db 3321 HEVDVNILDIPGAPEGLRHKIDITKESVVLKWDPELDDCGSPITNYVVEKQEDGGRWVPC 3380
QY 71 -----ANLTS----- 75
Db 3381 GETSDTSLKVNKLSGHEGHEKFRVAVNRQOTSAPLTSDBHAI VAKNPFDEPDAPTDTVPD 3440
QY 76 ----- 75
Db 3441 WDKDHVDLEKWPANDGGAPIDAYIVEKKDKFGDWECARVDGKTKTATADNLTPGETYO 3500
QY 76 ----- 75
Db 3501 FRVAVNKAAGPKSPDPTGNVAVAKPRRMAPKLNAGLLDLRIKAGTPIKLDIAFEGEPAP 3560
QY 76 -----SL 77
Db 3561 VAKWKANDATIDTGARADVNTPTSSAIIHFSAVRGDTGVYKIIIVENEHKGDTAQCNTV 3620
QY 78 LSVCG----- 82
Db 3621 LDVPGTPEGPKIDEIHKEGCTLNWKPPDNGGTDVLHYIVEKMDTSRGTWQEVGTFPDC 3680
QY 83 ----- 82
Db 3681 TAKVNKLVPGEYAFRVAVNLOGESKPLEAEPIIAKQDFVDPDPVDKPEVTDWDKDRI 3740
QY 83 -----WSOTIN----- 88
Db 3741 DIKWNPTANNAGPVTGYIIVEKKEGSAIWEAGTKPGTTFSSADNLKPGVEYFRVAVN 3800
QY 89 -----PEDDTP----- 95
Db 3801 AAGSPDSDPTDPOITKARYLKPILKTASRKIKIKAGFTNLEVDFIGADPTATWTVD 3860
QY 96 ----- 95
Db 3861 SGAALAPELLVDAKSSTTSITFFPSAKRADSGNYKLVKNELGEDEAIFEVIQDRPSAPE 3920
QY 96 ----- 95
Db 3921 GPLEVSQVTKDSCVLNWKPPKDDGAEISNYVVEKRDTKTNTWVPVSAFVTGTSITVPKL 3980
QY 96 ----- 95
Db 3981 TEGHEVEFRVMAENTFGRSDSLNTDEPVLAKDPFGTGPGRPEIVDTDNHDHIDIKWDPP 4040
```

Qy 96 -----DGNROVRG----- 120
Db 5121 TITWMAFLEDOGSKITGVNVEIREYSTLTWTVASDYNVREPEFTVDKLEFNDIEFRVVA 5180
Qy 121 ----- 120
Db 5181 INAAKGIPSLPSGPIKIQESGSGRPQIIVKVPEDTAQYNRRVFTCEAVGRPERTARWL 5240
Qy 121 ----- 120
Db 5241 RRGRELPESSRYRFEASDGVYKFTIKEVMDIDAGEYTVESNPNYSGSDTATANLVVQAPPV 5300
Qy 121 ----- 120
Db 5301 IEKDVNTIILPSGDLVRLKIYFSGTAPRHSVLNRREIDMDHPTIRIVEFDDHILITIP 5360
Qy 121 ----- 120
Db 5361 ALSVREAGRYEYTVSNDSEATTGFMNLNVTGLPEAPOGLHISNIGPSTATLSWRPPTVD 5420
Qy 121 ----- 120
Db 5421 GGSKITSVVEKRDLSKDEWVTVTSNVKDMNYIVTGLFENHEYFRVSAQENIGAPLV 5480
Qy 121 ----- 126
Db 5481 SEHPHIIARLPDPPTSPLENLEIVQVGGDYVTLQWOPRLSDGGRLRGYIVVEKQEEHDEM 5540
Qy 127 ----- 126
Db 5541 PRQNPNPPNNVNPVNLIDGRKYRVYFAVNDAGLSLAELOTLFOASGSGEGPKIVS 5600
Qy 127 ----- 126
Db 5601 PLSDLNEBVCRTFECEISGSPREYRWFPGCKELVDTSKYTLINKGDKOVLINDLTS 5660
Qy 127 ----- 126
Db 5661 DDDEYTCRATNSSGSTRANLRIKTPRVFIPPKYHGGYEAOKGETIELKIPYKAYPO 5720
Qy 127 ----- 126
Db 5721 GEARWTKDGEKIENSKFSITDDKATLRIISNASREDYGEYRVVENSUGSDSGTVNVT 5780
Qy 127 ----- 126
Db 5781 VADVPEPPRFPPIENILDEAVILSWKPPALDGGSLVTNYTIEKREAMGSGSWSPCAKRYT 5840
Qy 127 ----- 131
Db 5841 YTTIEGLRAGKQYEFRIIAENKHGQSKPCPTAPVLIPGDERKRRRGYVDVDEQGIVRGK 5900
Qy 132 ----- 131
Db 5901 GTVSSNYDNYVDIWKQYVQPVVEIKHDVLDHYDIHEELGTGAFGVVHRVTERATGNF 5960
Qy 132 ----- 131
Db 5961 AAKFVMTPHESKETVRKEIQTMSVLRHPTLVNLJHDAFEDDNEMWMIYEFMSGGELFEKV 6020
Qy 132 ----- 131
Db 6021 ADEHNKSEDEAVYMRQVCKGLCHMHENNYVHLDLKPENIMFTTKRSNELKLIDFGLTA 6080
Qy 132 ----- 133
Db 6081 HLDPKQSVKVTGTGAFAAPEVAEGKPVGYTDMKSVGLSYLLSGLSPFGENDDETL 6140
Qy 134 ----- 133
Db 6141 RNVKSCDNMDDSAFSGISEDGKDFIRKLLADPNTRMTIHOALEHPMLTFCNAPGRDSQ 6200
Qy 134 ----- 133

Qy 96 ----- 95
Db 4041 RDNGSPVDHYIERKDAKGRWIKWNTSPVQGTAFSTRVQKGHTYBYRVVAVNKAQPG 4100
Qy 96 ----- 95
Db 4101 QPSDSSAAATAKPMHEAPKFDLDDGKEFRVKAGEPLVITIPFTASPOPDISWTKGEGKP 4160
Qy 96 ----- 95
Db 4161 LAGVETDTSQTKLVIPSTRSDSGPVKIKAVNPYGEAEANIKITVIDKPGAPENITYPAV 4220
Qy 96 ----- 95
Db 4221 SRHTCTLNWDAPKDDGAEIAGYKIEYQEVGSIWDKVPLISGTAYTVRGLHGQQYRF 4280
Qy 96 ----- 95
Db 4281 RIRAEVGLSDYCOGCPVPIKDPDPGAPSTPEITGYDTNOVSLAWNPPRDDGGSPI 4340
Qy 96 ----- 97
Db 4341 GYVVERFEKGGDWAPVKMPVKCTECIVPGLHENETYQFRVRAVNAAGHGEPNSGSEP 4400
Qy 98 ----- 97
Db 4401 VTCRPYVEKPGADPRVGKITKNSAELTNRLRDGGAPIDGYIVEKKLGDNDWTRCN 4460
Qy 98 ----- 102
Db 4461 DKPVRDTAFEVNKLGEKEEYEFVRVIAVNSAGEGEPKPSDLVLEEQGRPIFDINNLIK 4520
Qy 103 ----- 102
Db 4521 ITVRAGETIQIRIPYAGGNPKPIIDLFGNSPIFENERTVVDVNPGEIIVITTGSKRSDA 4580
Qy 103 ----- 102
Db 4581 GPKISATNKYKDKTKLVFLVDAPKPTGPIRATDIQADAMTSLWRPPKDNGGDAITN 4640
Qy 103 ----- 102
Db 4641 YVVEKRTGGDWTVGVHVGTTTLVRNLDANTPYEFRVRAENQYGVGEPLETDDDAIVAKN 4700
Qy 103 ----- 102
Db 4701 PFDTPGAPQPEAVETSEAITLQWTRPTSDGAPIQGVVIEKREVSGSTEMTKAAGFNIL 4760
Qy 103 ----- 102
Db 4761 DTGHRVTGLTPKTYEFRVAAYNAAOGGEYSVNSVPITADNAPTRPKINMGMLTRDILAY 4820
Qy 103 ----- 102
Db 4821 AGERAKILVPPAASPAPKVFTSKGENKISPTDPRVKVEYSDFLATLTIEKSELTDGGLYF 4880
Qy 103 ----- 102
Db 4881 VELENSOGSDSASIRLVVDKPPASPOHIRVEDIAPDCCTLYMPPSSDGGSPITNYIVEK 4940
Qy 103 ----- 108
Db 4941 LDLRHSOGKEKVSFVRNLNVTGGLIKDNRYRFRVRAETOYGVSEPCELADVVAKYQ 5000
Qy 109 LELP----- 112
Db 5001 FEVPNQPEAPTVRDKDSTWAELEWDPDRDGGSKIIGYQVQYRDTSSGRWINAKMDLSEQC 5060
Qy 113 ----- 112
Db 5061 HARVTGLRQNGEFEFRIIAKNAAGFSKPSPSERCOLKSRFGPPGPIHVGAKSIGRNNHC 5120

Db 6201 IPSRVTYKIRDSIKTKYDANPELPLGLGRISNYSSLKRLKRPQEQYSIRDAFWDRSEAOQPRF 6260
QY 134 ----- 133
Db 6261 IVKPYGTGEGQSANFYCRVIASSPPVTVHKKDDRELKQSVKYMRYNGNDYGLTINRV 6320
QY 134 ----- 133
Db 6321 KGDDKGEYTVRAKNXSGYTKBEIVFLNVTRHSEPLKPEPLEPMKKAPSPRVVEFKERRSA 6380
QY 134 -----CLIT----- 137
Db 6381 PFTFHLNRLIKNHCKLTKCSLQGNPNPTIEMKDGHPVDEDRVQVSRSGVCSLEIF 6440
QY 138 -----EDTGF-----DLGVTI 148
Db 6441 NARVDDAGTVTVATNDLGDV 6462
RESULT 14
S57242
twitchin [similarity] - Caenorhabditis elegans
N:Alternate names: myosin-regulating protein
N:Contains: protein kinase (EC 2.7.1.-)
C:Species: Caenorhabditis elegans
C:Date: 28-Oct-1995 #sequence, revision 24-Oct-1997 #text, change 20-Jun-2000
A:Accession: S57242; S07571; S06797; S57218; T27934; T28030
R:Benian, G.M.; L'Hernault, S.W.; Morris, M.E.
submitted to the EMBL Data Library, February 1993
A:Description: Additional sequence complexity within twitching of Caenorhabditis elegans
A:Reference number: S57242
A:Accession: S57242
A:Molecule type: DNA
A:Residues: 1-6839 <BEN1>
A:Cross-references: EMBL: L10351
A:Experimental source: var. Bristol
R:Benian, G.
submitted to the EMBL Data Library, November 1989
A:Reference number: S07571
A:Accession: S07571
A:Molecule type: DNA
A:Residues: 792-6839 <BEN2>
A:Cross-references: EMBL: X15423; NID: G6897; PIDN: CAA33463.1; PID: G6898
A:Experimental source: var. Bristol
R:Benian, G.M.; Kiff, J.E.; Neckelmann, N.; Moerman, D.G.; Waterston, R.H.
Nature 342, 45-50, 1989
A:Title: Sequence of an unusually large protein implicated in regulation of myosin activity
A:Reference number: S06797; MUID: 90044042; PMID: 2812002
A:Accession: S06797
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 806-1175; 1178-1998, 'Y', 2000-3040, 'I', 3042-3335, 'I', 3337-5693; 5696-6359, 'I', 6360-6839
A:Cross-references: EMBL: X15423
A:Experimental source: var. Bristol
R:Benian, G.M.; L'Hernault, S.W.; Morris, M.E.
Genetics 134, 1097-1104, 1993
A:Title: Additional sequence complexity in the muscle gene, unc-22, and its encoded protein
A:Reference number: S57218; MUID: 93387664; PMID: 83971135
A:Accession: S57218
A:Molecule type: DNA
A:Residues: 2-99; 108-194, 'Q', 196-206; 374-468; 658-753 <BEN4>
A:Experimental source: var. Bristol
R:White, S.
submitted to the EMBL Data Library, May 1996
A:Reference number: 220442
A:Accession: T27934
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 'MGIPGKKCKQ', 19-6839 <W12>
A:Cross-references: EMBL: 273897; PIDN: CAA98064.1; GSPDB: GN00022; CESP: 2K617.1a
A:Experimental source: clone ZK617
R:Harrie, B.

submitted to the EMBL Data Library, May 1996
A:Reference number: 220458
A:Accession: T28030
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 'MGIPGKKCKQ', 19-6839 <W12>
A:Cross-references: EMBL: 273899; PIDN: CAA98081.1; GSPDB: GN00022; CESP: 2K617.1a
A:Experimental source: clone ZK829
C:Comment: Lack of unc-22 leads to a constant twitching of the body muscles.
C:Genetics:
A:Gene: unc-22; CESP: 2K617.1a
A:Map position: 4
A:Introns: 18/3; 69/3; 143/2; 176/3; 264/2; 387/3; 413/2; 471/1; 516/3; 550/3; 582/3;
152/3; 6691/3; 6776/1; 6808/3
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology
C:Keywords: ATP; autophosphorylation; duplication; muscle; phosphotransferase; serine
F1806-898, 899-990, 991-1083, 1084-1175, 1178-1273, 1474-1567, 1770-1864, 2066-2158, 2358-241
96-5790, 6263-6356, 6386-6478, 6541-6635, 6649-6742, 6745-6838/Region: motif 2
F1274-1372, 1373-1473, 1568-1670, 1671-1769, 1865-1964, 1965-2065, 2159-2258, 2259-2357, 241
23, 4215-4313, 4314-4415, 4416-4516, 4612-4710, 4711-4811, 4908-5009, 5010-5109, 5110-5210, 51
F:5940-6197/Domain: protein kinase homology <KIN>
F:5948-5956/Region: protein kinase ATP-binding motif
F:5971/Active site: Lys Metastatus predicted
Query Match 28.6%; Score 221; DB 2; Length 6839;
Best Local Similarity 1.3%; Pred. NO. 1.7e+02;
Matches 81; Conservative 28; Mismatches 32; Indels 6181; Gaps 25;
QY 3 GGILHL----- 8
Db 154 GGLYHAIFSDLDGDTYLCQLEIRGSSDAGQYRCNIRNDOGETNANLANFEEPDPSE 213
QY 9 ----- 8
Db 214 QERKRSTASPRSSRCGSRPSPKSKMSRECTPKRTLKPREGSPSKLRSRTSTPVNE 273
QY 9 ----- 8
Db 274 EYVSQSESRSSRTDKMEVDQVSGASKRPDGLPPPGGDEKLRAGSPSTRKSPRSKAS 333
QY 9 ----- 8
Db 334 TPRSCKSSAGGAAGTTGASATSATSGSASSASDRDKYTRPPIVLEASRSQTRIGG 393
QY 9 ----- 8
Db 394 SVVLEVOMQCHSSTIIEWYRDGTLVRNSSEYSQSFNSTAKLVNKLTEKSGLYKCHAK 453
QY 9 ----- 8
Db 454 CDYGEQSSAMVKIEQSDVEELMKHRKDAEDYQKEQKSTQLAETKKRVARRSKSKS 513
QY 9 -----ELL 11
Db 514 KSPAQAKKSTTSSEGRQEAHEVHKRSSVRPDPDEESQDLEIPSSGLTIPERRRELL 573
QY 12 VAVG----- 15
Db 574 GQVGESEDSEVSISBELPSFAGKPRKTDKPKKVSIAPIVSTNKSDDDEFTPRRSSI 633
QY 16 ----- 15
Db 634 DMRRESVQETLEKSTPLVPSGASGAPKIVEPVENVTVVENETAILTKCVSGSPAPTF 693
QY 16 ----- 15
Db 694 WFKGSRVISGGRFKHITDGKHTVALALLKRSQDEGPTYTLTIENVHGTSDADVKLLVT 753
QY 16 ----- 15
Db 754 SDNGLDFRAMLKHRSOAGFOKDGEGGAGGGGEGKKPMTEARRSOLFPGKVEKWDIP 813
QY 16 -----PD 17

Db	814	LPEKTVQQVQDKICEWKCTYSRPNKIRWYKDKRKEIFSGGLKXIVIKENKVTLLIINPE	873
Qy	18	VFOAHOEDTERY-----VLT-----	32
Db	874	V-----DQTKYTCENGVPHTHAQUTVLEPPMKYSFLNPLNTOEIIYRTKQAVLTCKVNT	928
Qy	33	-----	32
Db	929	PRAPLVMYRGSKAIOEGDPRFIIKDXAGRCRTLTIKEVEDDOAEWTARITQDVFSKVQV	988
Qy	33	-----	32
Db	989	YVEEPRHTFVPMKSOKVNESDLATLETVDNDKDAEVMVHDKRIDIDGVKFKVSSNR	1048
Qy	33	-----	32
Db	1049	KRRLIINGARIEDHGEYKCTTKDDRTMAQLIVDAKNKFIVALKDETVIEKDDVTLMCQTK	1108
Qy	33	-----NLNIG-----	37
Db	1109	DTKTPGIWFRNGKOISSMPGKFETQSRNGHTHLKIGKIEMNEADVVEIDQAGLRGCNV	1168
Qy	38	-----AELLR-----	42
Db	1169	TVLEAEKRPILNWKPKIEBAKGEPCVKVPQIHKGTGRGDPKAOILKNGKPIDEWRKL	1228
Qy	43	-----	42
Db	1229	VEVIIKDDVAEIVFKNPQLADTKWALELGNAGTALAPPFLVKDKPKPKGPLETKNV	1288
Qy	43	-----DPSLGA-----	48
Db	1289	TAEGLDLVMGTPDPDEGAPVKAVIIEMQESRGNNAKVGKTGDFVKOLKEHGEYKFR	1348
Qy	49	-----	48
Db	1349	VKALNECGLSPLTGESVLAKNPYGVPGPKNMDAIDVDKDHCTLAWEPPEDGGAPITG	1408
Qy	49	-----	48
Db	1409	YIIERRESEKDMHQVQTKPDCCELTDKKVVEDKEYLYRVKAVNKAGPGDPCDHGKPIK	1468
Qy	49	-----	48
Db	1469	MAKKASPEFTGGIKNLRLKVGETIKYDVPISGEPLPECLWVNGKPLKAVGRVKMSSE	1528
Qy	49	-----	48
Db	1529	RKXHMKIENAVRADSGKFTITLKNSSGCDSTATVTVVGRPTPKGPLDIADVCADGAT	1588
Qy	49	-----	48
Db	1589	LSWNPDDGGDPLTGYVEAQDMNDKGYIEVGKVDPNNTTTLKVLNGLRNKNGYKFRVKA	1648
Qy	49	-----	48
Db	1649	VNNEGESEPLSAQYTOIKDPWDEPKGRPEITDFDADRIDIAWEPHKOGGAPIEYI	1708
Qy	49	-----	48
Db	1709	VEVRDPOTKEWKVRYPDITNASISGLKEGKEYQFRVRAVNKAGPQSPSEKQLAKPK	1768
Qy	49	-----	48
Db	1769	FIPAWLKHDNLKSITVKAGATVRWEVKIGGEPIPEVKWFKGNOQLENGIQLTIDTRKNEH	1828
Qy	49	-----	48
Db	1829	TILCIPSMRSDVGEYBLTVKNSHGADEEKANLTVLDRPSKNGPLEVSDVPEDNLNSW	1888
Qy	49	-----	48

Db	1889	KPDDDDGGEPIEYVEVEKLDLTATGRWVPCAKVKOTKAHIDGLKKGQTYQFRVKAVNKSGA	1944
Qy	49	-----	48
Db	1949	SDALSTDKDTAKNPYDEPGKTGTPDVVDMDADRVSLEWEPKSGGAPITQYVIEKKGK	2008
Qy	49	-----QFRV-----	52
Db	2009	HGRDWCEGKVSDDOTNAEILGLKEGEYOFRVKAVNKAGGEASDPKRKVVAKPRNLKP	2068
Qy	53	-----	52
Db	2069	WIDREAMKTIITIKVGNDFEFDVVRGEBPPPKKEWIFNEKPYDDOKIRIESEDYKTRFVL	2128
Qy	53	-----HLVKMVL-----	60
Db	2129	GATRKHAGLYTLTATNASGSDKHSVEVIVLKGKPSPLGLEVSNNVEDRADLEWKVPEDD	2188
Qy	61	-----	60
Db	2189	GGADHYEIEKMDLATGRWPCGRSETTKTVPNLQPCHEYKFRVAVNKEGESDPLTT	2248
Qy	61	-----	60
Db	2249	NTAILAKNPYEVGKVDKPELVWDKDHVDLAWNAPDDGGAPIEAFIEKKDKNGRWEBA	2308
Qy	61	-----	60
Db	2309	LVPDQDKTATVPNLKEGEEOFRISARNKAGTGDPSDRVAVAKPRNLAPRIHREDLS	2368
Qy	61	-----	60
Db	2369	DTTVKVGATLKFIVHIDGEPADVTWSFNGKIGESKAQIENEPYISRPALPKALKOSG	2428
Qy	61	-----TEPEG-----	65
Db	2429	KYTIATNINGTDSVTINIKVKSPTKPKGPIEVDVPEDRATLQWKPPDDGGEPIDFY	2488
Qy	66	-----	65
Db	2489	EIEKNMTKDGIVPCGRSGDTHFTVDSLNGKDHKFRVKAVNSEGPSDPLETETDILAKN	2548
Qy	66	-----	65
Db	2549	PFDRPDRPCRPEPTDWDSDHDLKWDPLSDGGAPIEEOIEKRTKYGRWEPAITVPGQ	2608
Qy	66	-----	65
Db	2609	TTATVPDLTPNEEYFRVAVNKGSPSPSDASKAVIAKPRNLKPHIDRDALKNLTIKAG	2668
Qy	66	-----	65
Db	2669	QSISEFVPSGEPATVTWHFDPDREIRNGRVRKLDNPEYQSKLVVKQMGDSGTTIK	2728
Qy	66	-----	65
Db	2729	AVNNGEDEATVKINVIDKPTSPNGPLDVSVDVGHGHTVLNWRAPDDGGIPIENYVIEKY	2788
Qy	66	-----	65
Db	2789	DTASGRWVPAKVAGDKTTAVVDGLIPGHEYKFRVAAVNAEGESDPLETFTGLTKDPPD	2848
Qy	66	-----APNIT-----	70
Db	2849	KPGKTNAPETDWDKDHVLEWKPPANDGGAPIEYVVENKDEFSPFWNDAVHVPAGQTN	2908
Qy	71	-----	70
Db	2909	ATVGNLKEGSKYEFRIRAKNKAAGLGDPSDSASAVAKARNVPPVIDRNSIOEIKVKAGQDF	2968
Qy	71	-----	70
Db	2969	SLNIPVSGEPTTITWTFEGTVPVESDDRMKLNNEGDGKTFHVKRALRSDGTGTVIIKAENE	3028

```
QY 71 ----- 70
Db 3029 NGTDTAEVKTVLDFHPSRGLDVNTIVKDCGLAWKEPEDDGGAEISHYVIEKQDAAT 3088
QY 71 ----- 70
Db 3089 GRWTAGESKDTNFHVDLTLQGHEYKFRVKAVNRHGSDPLEAREAIIAKDPFDRADKPG 3148
QY 71 ----- 70
Db 3149 TPEIVDMKDHADLKWTPPADGGAPLEGYLVEMRTPSGDWVPAVTVGAGELTATVDGLK 3208
QY 71 ----- 70
Db 3209 PGQTYOFRKALNKAGESTPDPRMTVAKPRHLAKINRDMFVAORVAKAGOTLNFVNV 3268
QY 71 ----- 70
Db 3269 EGBAPAKIEWFLNGSPSSGGNTHIDNNTDNNTKLTTKSTARADSGYKIVATNESGKDE 3328
QY 71 ----- 70
Db 3329 HEVDVNILDIPGAPEGLRHKOITKESVVLKWDPELDDGSPITNYVVEKQEDGGRWVPC 3388
QY 71 ----- 75
Db 3389 GETSDTSLKVNKLSEGHEYKFRVKAVNRQGTSAPLTSDHAIVAKNPFDEPDAPTDTVPD 3448
QY 76 ----- 75
Db 3449 WDKDHVDLEWKPANDGGAPIDAYIVEKKDKFGDWECARVDGKTTKATADNLTGQETQ 3508
QY 76 ----- 75
Db 3509 FRVKAVNKGPGKSDPTGNVAVAKPRRMAPKLNLAGLLDLRIKAGTPIKLDIAFEGEPAP 3568
QY 76 ----- 82
Db 3569 VAKWKANDATIDTGARADVNTNTSSAIHIFSARVGDGTGVYKIIIVENEHGKTAQCQNVV 3628
QY 78 LSVCG ----- 82
Db 3629 LDVPGTPEGPKIDEIHKEGCTLNWKPPTDNGGTDVLHYIVEKMDTSRGTVQGVGTFPDC 3688
QY 83 ----- 82
Db 3689 TAKVNKLVEGKEYAFRVKAVNLOGESKPLEAEPPIIAKNQFVDPDPVKPEVTDMDKRI 3748
QY 83 ----- 88
Db 3749 DIKWNPTANNGGAPVTGYIVEKKEGSAIWEAGKTPGTTFSADNLKPGVEYEFRIAVN 3808
QY 89 ----- 95
Db 3809 AAGSPDSDPTDQITKARVLPKILTASRKIKIKAGFTHNLEVDFIGAPDPTATWTVDG 3868
QY 96 ----- 95
Db 3869 SGAALAPELLVDAKSSTTSIFFPSAKRADSGNYKLVKNELGEDEAIFEVIVQDRPSAPE 3928
QY 96 ----- 95
Db 3929 GPLEVSDVTKDSCVLNWKPKPDGGAEISNYVVEKRDTKTNTWVPVSASFVTGTSITVPKL 3988
QY 96 ----- 95
Db 3989 TEGHEYEFRVMAENTFGRSDSLNTEPVLAKDPFGTGPGRPEIIVDTDNHDHIDIKWDDP 4048
QY 96 ----- 95
Db 4049 RDNGGSPVDHYDIERKDAKTGRWIKVNTSPVQGTAFSDTRVQKHTYEVYRVAVNKGPG 4108
```

```
QY 96 ----- 95
Db 4109 QPSDSSAAATAKPMHEAPKFDLDLGKEFRVKGAGPLVITIPFTASPOQDISWTKEGGP 4168
QY 96 ----- 95
Db 4169 LAGVETTSQTKLVIPSTRRSSDSPVKIKAVNPYGEAEANIKITVIDKGPANENITYPAV 4228
QY 96 ----- 95
Db 4229 SRHTCTLNWDAPKDDGGAIEAGYKIEYQEVGSOIWDKVPCLISGTAYTVRGLEHGOQYRF 4288
QY 96 ----- 95
Db 4289 RIRAENAVGLSDYCOGVPVVIKDPFPPGAPSTPEITGYDTNOVSLAMNPPRDGGSPIL 4348
QY 96 ----- 97
Db 4349 GYVVERFEKRGGDWAPVMPVWKGTECIVPGLHENETVQFRVRAVNAAGHGEPSNGSEP 4408
QY 98 ----- 97
Db 4409 VTCRPYVEKPGADAPRVCKITKNSAELTNWRLDRDGGAPIDGYIVEKKKLGDNDRCN 4468
QY 98 ----- 102
Db 4469 DKPVRDTAFEVKNLGEKEBEYFRVIAVNSAGEPSKPSDLVLIBEQGRPIFDINNLDK 4528
QY 103 ----- 102
Db 4529 ITRAGETIQIRIPYAGNPKPIIDLFNNGSPIFENERTVVDVNPGEIVITTTGSKRSDA 4588
QY 103 ----- 102
Db 4589 GPKISATNKYCKDTCKLVFLVDAPGKPTGPIRATDIOADMTLSWRPPKONGDAITN 4648
QY 103 ----- 102
Db 4649 YVVEKRTPGDWMVTGVHVPVGTTLRVRLDANTPYEFRAENQYGVGEPLETDDAIVAKN 4708
QY 103 ----- 102
Db 4709 PFDTPGAPQPEAVETSEAITLOWTRPTSDGAPIQGYVIEKREVSTEWTKAAGNIL 4768
QY 103 ----- 102
Db 4769 DTKHRVTGLTPKTYEFRAVYNAAGQGEYSVNSVPIADNAPTRPKINMGMLTRDILAY 4828
QY 103 ----- 102
Db 4829 AGERAKILVPPFAASPAPKVTFSKGENKISPTDPRVKVEYSDFLATLTIEKSELTDGGLYF 4888
QY 103 ----- 102
Db 4889 VELENSQGSDSASIRLKVVDKASPQHIRVEDIAPDCCTLYMPPSSDGGSPITNYVBK 4948
QY 103 ----- 112
Db 4949 LDLRSDGKWEKVSFVRNLNTVVGGLIKDNRYRFRVRAETOYGVSEPCELADVVAKYQ 5008
QY 109 LELP ----- 112
Db 5009 FEVPNOPEAPTVRDKDSTWAELEMDPPRDGSGKIIGYQVQYRDTSSGRWINAKMDLSEOC 5068
QY 113 ----- 112
Db 5069 HARVTGLRONGEFEPRIIAKNAAGFKSPSPSERCOLKSRFGPGPIIHVGAKSIGRNNHC 5128
QY 113 ----- 120
Db 5129 TITWMAPELDGSGKITGYNVIREYVSTLWTVASDVNVRPEFTVDKLRFNDEYFPRVVA 5188
QY 121 ----- 120
```

Db 5189 INAGKIPSLPSGPIKIQESGSRPQIVVKPEDTAQPNRRVFTCEAVGRPEPTARWL 5248
Qy 121 ----- 120
Db 5249 RRGELPSSRYRFEASGVYKFTIKEDWIDAGEYTVESNPGYSDTATANLVQAPPV 5308
Qy 121 ----- 120
Db 5309 IEKDVNTILPSGDLVRLKIYFSGTAPRHSVLNREIDMDHPTIRIVEFDDHILITIP 5368
Qy 121 ----- 120
Db 5369 ALSVREAGRYEYTVSNDSEATTGFWLNTGLPEAQGPLHSNIGPSTATLSWRPPTD 5428
Qy 121 ----- 120
Db 5429 GSKITSYVEKRLDSKDEWTVTSNVKDMNYIVTGLFENHEYFRVSAQNGIGAPLV 5488
Qy 121 -----VTQLGG----- 126
Db 5489 SEHPILARLPDPTSPUNLEIVQGGDYVTLWSORPLSDGGGRGLRGYIVKEQEEHDEW 5548
Qy 127 ----- 126
Db 5549 FRCNQNPNNYVNPVLIDGRKYRYRFAVNDAGLSLAELDQTLFQASGSGGPKIVS 5608
Qy 127 ----- 126
Db 5609 PLSLDNEVGRCVTFECEISGSPREYRWFKCKELVDTSKYTLINKGDKQVLIINDLS 5668
Qy 127 ----- 126
Db 5669 DDADEYTCRATNSSGTRSTRANLRKTKPRVFIIPKYHGVEAOKGETIELKIPYAPQ 5728
Qy 127 ----- 126
Db 5729 GEARWKDGEKIENSKFSITDDKFATLRISASREDYGEYRVVENSVDGSGTVNVT 5788
Qy 127 ----- 126
Db 5789 VADVPEPRPFIENILDEAVILSWKPPALDGGSLVTNTYTIKREAMGSGWSPCAKRYT 5848
Qy 127 -----ACSEP----- 131
Db 5849 YTTIEGLRAGKQYEFRIIAENKHGSKCPEPTAPVLIPGDBKRRRGYVDVDEQGIKVRGK 5908
Qy 132 ----- 131
Db 5909 GTVSSNYDNYVDIWKQYYPQVVEIKHDPVLDHVDIHEELGTGAFGVVHRVTERATGNF 5968
Qy 132 ----- 131
Db 5969 AAKFVMTPHSEDKETVRKEIQTMSVLRHPTLVNLHDAPEDDNNEMVMYEFMSGGELFEK 6028
Qy 132 ----- 131
Db 6029 ADEHNKMSDEAVEYMQVCKGLCHMHENNYVHLDPENIMFTTKRSNELKLIIDFGLTA 6088
Qy 132 -----WS----- 133
Db 6089 HLDPKQSVKVTGTGAEPAAPEVAEGKPVGYTDMWSVGVLSYLLSGLSLPGGENDDET 6148
Qy 134 ----- 133
Db 6149 RNVKSCDNMDDSAFGISGDKDFIRKLLADPNTRMTIHQALEHPWLTPGNAPGRDSQ 6208
Qy 134 ----- 133
Db 6209 IPSSRYTKIRDSIKTKYDAMPPEPLPLGRIGNYSSLRKXRPQEVSIIRDAFWRSEAPRF 6268
Qy 134 ----- 133

Db 6269 IVKPYGTGEGOSANFYCRVIASSPPVVTWTHKDDRELKQSVKMKRYNGNDYGLTINRV 6328
Qy 134 ----- 133
Db 6329 KGDDKGEYTVRAKNSVGTKEEIVFLNVRHSEPLKKEPLEPMKKAPSPRVEEFKERRSA 6388
Qy 134 -----CLIT----- 137
Db 6389 PPTFFHLNRLNIQKNHOCKLTCSLQGNPNPTIEWMKDGHVDEDRVQVSPRSGVCSLEIF 6448
Qy 138 -----EDTGF-----DLGVTI 148
Db 6449 NARVDDAGTYVTATNDLGDVDV 6470
RESULT 15
T30283
polyketide synthase - Streptomyces sp. (strain MA6548)
C:Species: Streptomyces sp.
A:Variety: strain MA6548
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C:Accession: T30283
R:Motamedi, H.; Cai, S.J.; Shafiee, S.J.; Elliston, K.O.
Eur. J. Biochem. 244, 74-80, 1997
A:Title: Structural organization of a multifunctional polyketide synthase involved in
A:Reference number: Z20806; MUID:97217427; PMID:9063448
A:Accession: T30283
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6420 <MOT>
A:Cross-references: EMBL:Y10438; NID:e1014806; PID:e290681; PIDN:CAA71463.1
C:Genetics:
A:Note: fkbA
C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier pr
C:Keywords: carrier protein
F:51-433/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F:1930-2325/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F:3391-3462/Domain: acyl carrier protein homology <ACP1>
F:3505-3900/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F:3983-4234/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F:5307-5378/Domain: acyl carrier protein homology <ACP2>
F:5431-5831/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F:5932-6206/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F:6293-6364/Domain: acyl carrier protein homology <ACP3>
Query Match 28.5%; Score 220; DB 2; Length 6420;
Best Local Similarity 1.5%; Pred. No. 1.5e+02;
Matches 88; Conservative 25; Mismatches 35; Indels 5843; Gaps 27;
Qy 1 AAGGI-----LHLE----- 9
Db 360 AAGAGIIMKVLAMRHGTLPTLHVDTSPSHVDWTTGDVELLTQARPWPQTNPRHAGI 419
Qy 10 -----LVAV----- 14
Db 420 SFEVSGTNAHVIESHPRTAPDGTSGSTHPVLLISARTPEALDEHTTIRAFLDNP 479
Qy 15 -----GPDV----- 18
Db 480 GADRVTVQTLARRTPPEHRAVLGLDTLITADPNAGSGPVVYVYSGOSTLHPHTGHOLAA 539
Qy 19 ----- 18
Db 540 TYSVFADWGEVLGHLNADOGPATHFAHQIALTALLRSWGITPHAVIGHSLGEISAACAA 599
Qy 19 ----- 18
Db 600 GVLSIGDASALLAARSRLMDELPTGGAMVTVLTSEENALRALRPGVEIAAVNGPHSVLS 659
Qy 19 ----- 18
Db 660 GDEGPVLDVAQQLGIHRLPTRHAGHSARMDPLVAPLLEAASGLTYHOPHTAIPEDPTTA 719

QY 19 -----FOAHQE----- 24
Db 720 AYWAROVROVRFQAAERYPGATFLEIGNQDLSPVVDGIPQTGTGPEVQALHTALAR 779
QY 25 ----- 24
Db 780 LHTRGGVDMPTVLGSDRAPVALPTYPFOHKDYWLRLATAQVDVTGAGQEKVAHPLLGAUV 839
QY 25 ----- 24
Db 840 ALPGTGTVLTRGRVSLASHPNLGGHAVNGTVLLPGAFLFLAVRAGDETGSDDLHELVE 899
QY 25 -----DTER----- 28
Db 900 TPLTLPATGSAISVEAEPDDTGRAVTVHARADGAALWTRHATGFLGKAATAETAETGD 959
QY 29 ----- 28
Db 960 PAPMPAGARQVDVYERFAGIGYAGSSFOGLLAWRDGTVHAVALPDEQVPDAA 1019
QY 29 ----- 28
Db 1020 RFTLHPALLDAFQAGALSALDAPGEAARLPFSFQDVRHAAGATRARVTVSRDGRSTV 1079
QY 29 ----- 28
Db 1080 RMTGEDQPVAEYGVSRPYAGSGSDDELLRPVMTLPMAPSAADPRVEVLDADPLD 1139
QY 29 ----- 28
Db 1140 GEPAAATRELTVARVLSLOHRLSATDDVTLVVRIGTGAAAAATAGLVRSAQENPGRFVL 1199
QY 29 ----- 28
Db 1200 VEAPDTPADLLTACAALDEPQLAVRDGVLFAPLVRMSDPVHGPLPLPDGWLHAHAS 1259
QY 29 ----- 28
Db 1260 GVLQDVTLAADGTPRRALEAGEVRVDVRAAGLNFRDVLIALGTGATALGGEAGVVE 1319
QY 29 -----YVLT 32
Db 1320 TGPGVADLAPGDRVFLGTRGGIGTAVTDRRWLARIPOGWTFTATAASVPIAFATAWGLV 1379
QY 33 NL-----NIGAE----- 40
Db 1380 DLAGLRAGEKVLVHAATGGVMAATQVARHLGAEVYATAGTGKQHVLRAGLPGTHVADS 1439
QY 41 -----LRDP----- 44
Db 1440 RTTAFRTDPPMDVVLNSLTGBFIDASLELLDADGRFVEMGRAELRDPATIAPAYLPFDL 1499
QY 45 -----SLGAQFR----- 51
Db 1500 LDAGTDRIGEILGELLRLFDAGALTPLPVRAWDIROARDALGYMSRARHIGKNVLTLP 1559
QY 52 -----VHL----- 54
Db 1560 LDPEGAIVITGSGTLAGILARHVSQAQHVYLLSRSAPECTPGIHLPCDVSRRQQAAL 1619
QY 55 ----- 54
Db 1620 RRIDRPITAVVHTAGALDDGTLPSTLPGRFDTLPRKADGAWHLHELTKDQDLAAFLVYS 1679
QY 55 ----- 54
Db 1680 SAAGVLGNAGOGNYAAANAFDALAEORRAAGLPALSIAWGLWEDASGLTAELSEADQR 1739
QY 55 ----- 54
Db 1740 MRRSGFTTARQGMRLYEARRTGSPPVVAAALDDAPDVPFLLRGURRTTVRRAAVRERS 1799
QY 55 ----- 54

Db 1800 LADSPCCPTTSAPTSPSRSSWNSTATVLGHLGAEDIPATTTFKELGIDSLTAVOLRNAL 1859
QY 55 ----- 54
Db 1860 TTATGVRLNATAVDFPTTPRALAARLGDELACTRAPVAARATAAANDEPLAIVGMACR 1919
QY 55 ----- 54
Db 1920 LPGVASPOELWRLVASGTDAITFEPPADRGWDVLDALYDPPDAIGKTFVRHGGFLDGTG 1979
QY 55 ----- 54
Db 1980 FDAFFGISPREALAMPQORVLLTSEAFESAGITPDAARSGDTGVFIGAFSYGYGTG 2039
QY 55 ----- 54
Db 2040 ADTNFGATGSGTSVLSGRLSYFYGLEGPSVTVDACSSSLVALHQAGOSLSRSGESLAL 2099
QY 55 ----- 54
Db 2100 VGGVTVMASPGGFVBFSRQGLAPDGRAKAFGAGADGTSFAEGAGALVVERLSDAERHGH 2159
QY 55 ----- 54
Db 2160 TVLALVRGSAANSDBGASNGLSAPNGPSQSERVIHQALANAKLTPADVDAVEAHGTGTRLGD 2219
QY 55 -----VKMV----- 58
Db 2220 PIEAQLLATYQDQDRATPLLLGSLKSNIGHAQAASGVAGIIMVQAIRHGELPPTLHADE 2279
QY 59 ----- 58
Db 2280 PSHVDWTAGAVELLTSARPWGTGRPRRAAVSSFGVSGTNAHIIIEAGPVKTGPVEAGA 2339
QY 59 ----- 58
Db 2340 IEAGVEVGPVEAGPLPAAPSPAGEDLPLLVSARSPEALDEQIGRLRAYLDTGPGVDRA 2399
QY 59 ----- 58
Db 2400 AVAQTLLARRHTHRAVLLGDTVIGAPPADQADELVFVYSGGTQHPAMGEOLAAAPVF 2459
QY 59 ----- 58
Db 2460 ADAMHDALRLDDPDHPDTPRSQHTLFAHQAAFTALLRSWDITPHAVIGHSLGEITAAVA 2519
QY 59 -----ILTE----- 62
Db 2520 AGILSLDDACTLITTRARLMHTLPPPGAMVTULTSEEARQALRPGVEIAWFGPHSVVL 2579
QY 63 ----- 62
Db 2580 SGDEDAVLDAORLGIHRLHRLPAPHAGSAHMEPVAAELLATTRELYRPHPTAIPNDPTT 2639
QY 63 ----- 62
Db 2640 AEYWAQVNRNPNVLFNAHTQRYPDVFEIGPGQDLSPLVDGIALQNGTADVEVHALHTALA 2699
QY 63 ----- 62
Db 2700 RLFTRGATLDWSRILGGASRHPDPVPSYAFQRRPYWIESAPPATADSGHPVLGTGVAVAG 2759
QY 63 ----- 62
Db 2760 SPGRVFTGVPAGADRAVPIAELALAAAADATCATVEQLDVTSPGSGSARGRATAQTWVD 2819
QY 63 ----- 62
Db 2820 EPAADGRRRFTVTRVGDAPWTLHAEGVLRPGRVPOPEAVDTAWPPPGAVPADGLPCAHR 2879
QY 63 ----- 62


```
Db 2880 RADQVFEAEVSDPDGFVAHPDLLDAVFSAVGDSRQQTGWRDLAVHASDATVLRACLTR 2939
Qy 63 ----- 62
Db 2940 RDSGVELAAPDAGMPVLTAESVTLGEVASAGSDESGLRLLEWLPVAEHAHDGADEL 2999
Qy 63 PEG----- 65
Db 3000 PEGYTLITATHPDDDDPTNPHTPTRHTQTTRVLTAQHHLITTNHTLIVHHTDDPPG 3059
Qy 66 ----- 65
Db 3060 AAVTGLTRTAQNEHPRIRHETIHHPTPLPLTQTLTLQHPLRLTNNTLHTPHLTPIIT 3119
Qy 66 ----- 65
Db 3120 HHNTTTTTPNPNHAILTGGSGTLAGILARHLNHPHTYLLSRTPPPPTPGTHIP 3179
Qy 66 ----- 65
Db 3180 COLDPTQITQALTHI PQPLTGI FHTAATLDDATLTNLTPQHLTTTLQPKADAAWHLHHH 3239
Qy 66 ----- 65
Db 3240 TQNOPLTHFVLYSSAAATLSPGOANYAANAFLDALATHRHTQCPATTIAGMWHTTT 3299
Qy 66 ----- 65
Db 3300 TLTSQLTSDRDRIRRGGLPISDDEGMRLYDAAVSGEDEFVLAAMDPQPMAGDVVPI 3359
Qy 66 ----- 65
Db 3360 LSGLRKSARTARTQOTFAORLAELPAADRDTALVTVLSDATAAVLGHADASGIAPTTTF 3419
Qy 66 ----- 65
Db 3420 KDLGIDSUTAIELNRNLAETGLRLSATMVFDPHPTPRVLAELRTDLFGTAAPLPARTAR 3479
Qy 66 ----- 65
Db 3480 TRHDEPLAIVGMACRLPGGVTS PEDMLRLVAGTDAITEFPTDRGWIDRMFDPDPDAPG 3539
Qy 66 ----- 65
Db 3540 KTYVRHGGFLSEAAAGFDAAPFGISPREAWAMPQORVILETVWEAFENAGIVPTLRGSD 3599
Qy 66 ----- 65
Db 3600 TGVFMGAFSHGYGAGDDLGGFGATATQNSVLSGRLSYFFGMEGPAVTIDTACSSSMVALH 3659
Qy 66 ----- 65
Db 3660 QAAQSLRDGCSLALAGGVTYMPTPLGYVEFCRQGLAPDGRAKAFAGADGTSFSEGAG 3719
Qy 66 ----- 68
Db 3720 VLVVERLSDAERNGHTVLALVRSSAVNQDGASISAPNGSQORVIRQALDKAGLAPAD 3779
Qy 69 ----- 68
Db 3780 VDVVEAHGTGTPLDPIEAQAI IATYGQDRDTPLYLGSVKSNIIGHTOTTAGLAGVIKVMV 3839
Qy 69 ----- 68
Db 3840 AMRHGLLPKTLHVDEPSSHVDWSAGAVELLTEARWPDPDRPRAGVSSLSIGSTNAHVI 3899
Qy 69 ----- 68
Db 3900 LEGVAESSVRSESSGLVPLPVSARTESSLALQVERLGEYVRGARDLAAVADGLVRGTV 3959
Qy 69 ----- 68
Db 3960 FGHRVAVLLGESTVAGVAGAESRTVFVFPQGQSQVMGMGRELMSGCEVFAARMRECAAVLE 4019
```

```
Qy 69 ----- 68
Db 4020 PYTGMDLLDVLGEAVVAERVELQPSAWAVASVLAALWQAHGVSPPDAVIGHSGOEIAAAC 4079
Qy 69 ----- 68
Db 4080 VAGALSLEDAARIVALRSQTIAAHLAAGRGAMASIALPATAVETVEGVWVAARNGPSTVV 4139
Qy 69 ----- ITANLTSSLLSVCCWS 84
Db 4140 AGDPSAVERVLARYEAEGVRVRRIAVDYASHTPHVEAIQEQALADVLGDIITSSAPSPWMS 4199
Qy 85 ----- 84
Db 4200 TVDGGWTEPAGDDYWRNLQRPVAMDTAIGELDGSFIECSAHPVLLPALDOERTVASL 4259
Qy 85 ----- 84
Db 4260 RTDDGGHDFELTALAQAQWTOGADVDMHTLLIAPAPDRLLDLPTYPFDHKRYWIEATGAADL 4319
Qy 85 ----- 84
Db 4320 TALGLTDTAHPMLAATAALPTNDGTVLTRGRISLRTHPWLADHTIQDTVLLPGTAFLELVM 4379
Qy 85 -----QTIN----- 88
Db 4380 RAGDEVCDDTIDELVETPLTLPVTGAVDLTVTVEQADATGHRPVSVHARPACTDWTWRH 4439
Qy 89 ----- 88
Db 4440 ATGTLGRTEDTAPDNTSNFPQWPPTGTALDVTEFYAQLAALGYRFVPTFRGLRAAMRAGD 4499
Qy 89 -----PED----- 91
Db 4500 TVYAEVALPEDRAADADHFGVHPALLDAALQSGSLLMLESDEGQVQLPFSWHGVRFHAT 4559
Qy 92 ----- 91
Db 4560 GATSLRAIVPGPDGURLRAADSENHPVATIDALVTRPPADLAPSDPVLRVGWTAVPVPA 4619
Qy 92 ----- 91
Db 4620 EVRPSDADVLMHSDADPIQETRDLTTRVLDALLTTEGTLIVQVTTGGLAAMAAAGLVRT 4679
Qy 92 -----DTPG----- 96
Db 4680 AQSEQPCRFVLVETDPGEVLDGARCDMIAALGEPRLRLRGNRFEARLVRAAAALTPPDT 4739
Qy 97 ----- 96
Db 4740 GSWQLRQSATGSLDDVAIVPTGEPDRPLGAGEVRIAVRAAGLNFRDRTVALGVVADRPL 4799
Qy 97 ----- 96
Db 4800 GSEAAGVILETGPVHDLAPGDRVLGMLAGAGCPVAITDORLLGRMPDGTWTFPQASVVT 4859
Qy 97 -----HADLVLYTR----- 106
Db 4860 VFATTWGLVDLAELRPGEKVLIHAAATGVGAAAQVQIARHLADADVVATATSTPKOHLDDLE 4919
Qy 107 -----F 107
Db 4920 PTHIADSRITAFADTFFPADVVVLSLTGPFLDASAQLLTPGGRFIEMGKTDIRHHVQOPF 4979
Qy 108 DLEL----- 111
Db 4980 DLMVAGPDRLQIIITQLLDLFANGTLHPLPVHAWDVRQAREAFGMSSGRHTCKLVLTVP 5039
Qy 112 -----PDG----- 114
Db 5040 RPLDPEGAVIITGGSGTLAGHLARHLDPHTYLLSRTPPPPTTTPGTHLPCDVSDPTQLAT 5099
```

```

QY 115 ----- 114
Db 5100 ALAQIPOPITGIFHTAATLDALLHNLTPORIDTVLOPKADAAMHLHHHTHATDLAAAF 5159
QY 115 ----- 120
Db 5160 VLYSAVAGMSPGOGNYAANAFDUAAEHQAMGLPAQSLANGMWADVSALTAKLTDA 5219
QY 121 ----- 126
Db 5220 DRGRIIRGGFPPLSAACDMRLFDSAASTRSSRRSPSDVTRLDGPVAPLLRGLVAHRSG 5279
QY 127 ----- 126
Db 5280 PVRAVNADDEPLAARLAGHTAAEQRRVMOEVLVLRQAAAVLAYGLGEQVAADRPRDLG 5339
QY 127 ----- 126
Db 5340 FDSLTAVDLNRNLAAETGLRLPTTVVFSHTAEALATHLLELLDAPALTTGAPLPVTA 5399
QY 127 ----- 126
Db 5400 APGTAARDQDEPIAIVAMACRLPGGVTSPEELWRLVSGTDAITTPGDRGWDLDALYDP 5459
QY 127 ----- 126
Db 5460 DPDAVGKAYSIRGGFLEGAEEFAAFFDISPRESLGMDDPQORLLELTAEWAIERGRINPA 5519
QY 127 ----- 129
Db 5520 SLHGREIGVYVGAAGGYGLCAEDTEGNATGGTSLLSGRLAYVLEGPVAVTVDTACS 5579
QY 130 ----- 129
Db 5580 SSLVALHLAGOGLRGECELALAGVSVLSSPAAFVFSRQRLAAGDRCKSPGSGADGT 5639
QY 130 ----- 129
Db 5640 TWAEQVGLVLEQLSDAERLHTVLAVVRSVAVTSDGASNGLTAPNGLAQORVIRKALAA 5699
QY 130 ----- 132
Db 5700 AGLTAADVLDLVEGHGTGTRLDGPVEADALLATYQNGRGPVWLGSLKSNIGHATAAGVA 5759
QY 133 ----- 132
Db 5760 GVIKTVOAIGAGTMPRTLHADESPAVDWTAGVSVLLTGNRPWMTSVPGERPSRSGSA 5819
QY 133 ----- 135
Db 5820 GRTLTSLSNSTVSPWRRDLARNPGRCLGCCPRPLPCGRRPPVCAIIRPSGPDPLD 5879
QY 136 ----- 135
Db 5880 IGYAQTSAQFTHRAAEAAASPDGFRAGLDSVADGAEPSSAEPPRSARRELFDGQGA 5939
QY 136 ----- 135
Db 5940 QRCVMGRELHGRFPVFAAAWDEVSDFGKLEHSPDVFHGEHGLAHDTLIAQVGLFTL 5999
QY 136 ----- 135
Db 6000 EVALLRLEHMGVRPDVVVGHSGVEVTAAYAAGVLTADATTLIVARGRALRPPGAMT 6059
QY 136 ----- 135
Db 6060 AVEGSPAECVCAFTDLDAIAVNGPSAVVLTGAPDDVAAPERWAAAGRRAKRLDVGHAFHS 6119
QY 136 ----- 135
Db 6120 RHDGALDDFRGVLESFAGAAALPVVSTTTTGRDAAGDLATPEHHLRHARRPVLYADAVR 6179
QY 136 ----- 135

```

```

Db 6180 ELADLGVMNFVAVGPGSALASAASSENTGGSAGTYHAVLRARTGEESAALTAVAEHLHAHA 6239
QY 136 ----- 135
Db 6240 PVDLAAVLAGRPVDLPVYPFOHRSYMLAPAVGGSPATAVPDTGRPTPEPDDLTVAEIV 6299
QY 136 ----- 148
Db 6300 RRRRAALLGIADPGVDADTTFFALGFDSLAVORLNRNOLTAATGLDLPTAV 6350

```

Search completed: March 20, 2003, 12:41:52
Job time : 204 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:35:42 ; Search time 12 Seconds
(without alignments)
511.541 Million cell updates/sec

Title: SEQID_1_15FUSED

Perfect score: 773

Sequence: 1 AAGGILHLELLVAAGPDVFQ.....SPTWSCLITETGFDLGVTI 148

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	237	30.7	997	1	AT57_HUMAN
2	239	29.6	1592	1	AT12_HUMAN
3	246	29.2	5065	1	EPPL_HUMAN
4	244	29.0	5255	1	BACA_BACLI
5	222	28.7	5217	1	HTS1_COCCA
6	216	27.9	4829	1	BIR6_HUMAN
7	215	27.8	4349	1	FAT2_HUMAN
8	212	27.4	5035	1	RYR1_PIG
9	211	27.3	4351	1	FAT2_RAT
10	210	27.2	5038	1	RYR1_HUMAN
11	210	27.2	6486	1	TYCC_BACBR
12	209	27.0	3649	1	ACVS_NOCLA
13	209	27.0	5147	1	FAT_DROME
14	208	26.9	4303	1	PKD1_HUMAN
15	207	26.8	4590	1	FATH_HUMAN
16	206	26.6	4969	1	RYR2_RABIT
17	205	26.5	4639	1	DYHC_DROME
18	204	26.4	890	1	AT58_HUMAN
19	204	26.4	5017	1	RYR1_RABIT
20	203	26.3	4967	1	RYR2_HUMAN
21	203	26.3	6359	1	BACC_BACLI
22	202	26.1	1629	1	AT59_HUMAN
23	202	26.1	3988	1	POLG_BVDVN
24	200	25.9	3859	1	RPOA_LELV
25	200	25.9	4870	1	RYR3_HUMAN
26	199	25.7	967	1	AT51_RAT
27	199	25.7	3898	1	POLG_HCVB
28	198	25.6	3396	1	PGCV_HUMAN
29	198	25.6	4128	1	PRXD_HUMAN
30	198	25.6	4540	1	DYHC_PARTE
31	198	25.6	5327	1	ACF7_MOUSE
32	197	25.5	968	1	AT51_MOUSE
33	196	25.4	860	1	AT56_HUMAN

34	196	25.4	4349	1	DYHC_FUSSO	P78716 fusarium so
35	196	25.4	4486	1	DYH9_HUMAN	O9HYC9 homo sapien
36	196	25.4	4655	1	LRP2_HUMAN	P93164 homo sapien
37	195	25.2	3354	1	CADN_HUMAN	O9N251 homo sapien
38	195	25.2	3712	1	ACVS_CEPAC	P25464 cephalospor
39	195	25.2	3898	1	POLG_HCVB	P19712 hog cholera
40	195	25.2	4092	1	DYHC_YEAST	P36022 saccharomyc
41	195	25.2	4128	1	PRKD_MOUSE	P97313 mus musculu
42	195	25.2	4568	1	DYHB_CHLRE	O39565 chlamydomon
43	194	25.1	837	1	AT54_HUMAN	O75173 homo sapien
44	194	25.1	4485	1	DYHG_CHLRE	O39575 chlamydomon
45	194	25.1	4644	1	DYHC_MOUSE	O9JHU4 mus musculu

ALIGNMENTS

RESULT 1

AT57_HUMAN

ID AT57_HUMAN STANDARD; PRT; 997 AA.

AC

O9UKP4;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase

DE with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7).

GN

ADAMTS7.

OS

Homo sapiens (Human).

OC

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX

NCBI_TaxID=9606;

RN

[1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9395124; PubMed=10464288;

RA Hurekainen T.L., Hirohata S., Seidin M.P., Apté S.S.;

RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, Novel Members of a New Family of

RT Zinc Metalloproteases.";

RL J. Biol. Chem. 274:25555-25563(1999).

CC -!- COPACATOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular

CC matrix (By similarity).

CC -!- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG,

CC LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS.

CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT

CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY

CC SIMILARITY).

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announcements>

CC or send an email to license@ebi-sib.ch).

CC -----

CC EMBL; AF140675; AAD56358.1; -

CC HSSP; P15167; IATL.

CC MEROPS; M12.231; -

CC Genew; HGNC:223; ADAMTS7.

CC MIM; 605009; -

CC InterPro; IPR001762; Disintegrin.

CC InterPro; IPR002870; Pep_M12B_propep.

CC InterPro; IPR001590; Reprolysin.

CC InterPro; IPR000884; TSP1.

CC InterPro; IPR000130; Zn_MTpeptidase.

CC Pfam; PF00090; tsp_1; 1

CC Pfam; PF01421; Reprolysin; 1.

CC Pfam; PF01582; Pep_M12B_propep; 1.

CC SMART; SM00209; TSP1; 2.

```

DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSPI; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 232 BY SIMILARITY.
FT CHAIN 233 997 ADAMTS-7.
FT SITE 204 204 CYSTEINE SWITCH (POTENTIAL).
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 389 389 BY SIMILARITY.
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 462 537 *DISINTEGRIN-LIKE.
FT DOMAIN 538 594 TSP TYPE-1 1.
FT DOMAIN 595 697 CYS-RICH.
FT DOMAIN 698 914 SPACER.
FT DOMAIN 915 990 TSP TYPE-1 2.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED02FC104 CRC64;

Query Match 30.7%; Score 237; DB 1; Length 997;
Best Local Similarity 18.2%; Pred. No. 0.026;
Matches 61; Conservative 24; Mismatches 62; Indels 188; Gaps 6;

OY 1 AAGGILHLEL-----LVAVG----- 15
DB 53 AGSFLSYELMPALRKRDVSRDAPAFYELQYRGRELNFNLTAQHLLAPGFVSEFR 112
OY 16 -----PD-----VFOAHEED----- 25
DB 113 RGGLGRAHTRANTPACHLLGEVDPELEGGLAAISACDGLKGVFOLSNEYFIEPLDSAP 172
OY 26 -----TERYVLTNLNIGAEILLRDPSPGAFRVHLV 55
DB 173 ARPGHAQPHVYKQAPERLAQRGDSAPSTCGVQVYPELESRRERWEQRQWRRLR 232
OY 26 -----TERYVLTNLNIGAEILLRDPSPGAFRVHLV 55
DB 233 LHQRSVSEKWKWETLVADAKMVEYHGVQVPEVSYVLTIMNVVAGLFHDPISGNPIHITV 292
OY 56 KMWILTEPEGAPNITANITSSLLSCVGSOTINPEDDTPGHADLVLTFRDLELPDGN 115
DB 293 RVLVLEDEEDLKITHADNTLKSKCKQKSNKMGDAHPHHTAILLTRKDL-CAAMN 351
OY 116 R--QVRGVTLGGACSPWSCLITEDTGFDLGVTI 148
DB 352 RPECETGLSHVAGMCQPHRSCSINEDTGLPLAFTV 386

RESULT 2
AT12_HUMAN STANDARD; PRT; 1593 AA.
AC PS0397;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-
DE TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=21264577; PubMed=11279086;
RA Cal S., Argueelles J.M., Fernandez P.L., Lopez-Otin C.;

```

```

RT "Identification, characterization, and intracellular processing of
RT ADAM-TS12, a novel human disintegrin with a complex structural
RT organization involving multiple thrombospondin-1 repeats.";
RL J. Biol. Chem. 276:17932-17940(2001).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
CC -1- PTM: IS SUBJECT TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
CC DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AJ250725; CAC20419.1; -
DR GenM; HGNC:14605; ADAMTS12.
DR MIM; 606184; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_MTPeptidse.
DR Pfam; PF00090; tsp_1; 6.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPI; 8.
DR PROSITE; PS00215; ADAM_MEPRO; 1. FALSE_NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS00092; TSPI; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 240 BY SIMILARITY.
FT CHAIN 241 1593 ADAMTS-12.
FT DOMAIN 465 544 DISINTEGRIN-LIKE.
FT DOMAIN 545 596 TSP TYPE-1 1.
FT DOMAIN 597 700 CYS-RICH.
FT DOMAIN 701 826 SPACER 1.
FT DOMAIN 827 881 TSP TYPE-1 2.
FT DOMAIN 886 943 TSP TYPE-1 3.
FT DOMAIN 947 995 TSP TYPE-1 4.
FT DOMAIN 996 1315 SPACER 2.
FT DOMAIN 1316 1364 TSP TYPE-1 5.
FT DOMAIN 1367 1423 TSP TYPE-1 6.
FT DOMAIN 1426 1471 TSP TYPE-1 7.
FT DOMAIN 1471 1471 TSP TYPE-1 8.
FT DOMAIN 302 305 POLY-GLU.
FT SITE 208 208 CYSTEINE SWITCH (POTENTIAL).
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 393 393 BY SIMILARITY.
FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 685 685 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 951 951 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1275 1275 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1300 1300 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1320 1320 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1503 1503 N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ SEQUENCE 1593 AA; 177545 MW; 07F9F48E63B83A3 CRC64;

Query Match 29.6%; Score 229; DB 1; Length 1593;
Best Local Similarity 18.3%; Pred. No. 0.26;
Matches 49; Conservative 35; Mismatches 61; Indels 123; Gaps 4;

Oy 3 GGILHLELLVAVGP----- 16
Db 124 GNLSHVKNMASSAPLCHLSGTVLQGGTRVGTAAALSACHGLTGFFQLPHGDFFTIEPVKKHP 183
Oy 17 -----DVFOAH----- 22
Db 184 LVGGYHPHIVYRQKVPETKCTGLKDSVNISSQOELWREKWRHNLPSRSLRSRSIS 243
Oy 23 -----QEDTERVLTNLTNIGAEILLRDPGLGQAFRVHLVGMVLT 62
Db 244 KERWETLVVADTKMIEYHGSNVSILTIMNMTGLPHNPSIGNAIIHVVRLLLEE 303
Oy 63 PEGAPNITANLTSSLSVCGMSOTINPEDTDPGHADLVLYITRFOLELPDGNR--QVRG 120
Db 304 EEOGLKIVHAETLSFCMKQKSIKPKGDLNVPVHDVAVLLTRKDI-CAGFNRPCTELG 362
Oy 121 VTOLGSGCPTWSCLITETDGTGFDLVGTI 148
Db 363 LSHLSGMCQPHRSNCINEDSGLPLFTI 390

RESULT 3.
EPPL HUMAN
ID EPPL HUMAN STANDARD; PRT; 5065 AA.
AC P58107;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epiplakin (450 kDa epidermal antigen).
GN EPIPK1 OR EPIPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=21201183; PubMed=11278896;
RA Fujiwara S., Takeo N., Otani Y., Parry D.A.D., Kunimatsu M., Lu R.,
RA Sasaki M., Matsuo N., Khaleduzzaman M., Yoshioka H.;
RT "Epiplakin, a novel member of the plakin family originally identified
RT as a 450-kDa human epidermal autoantigen: structure and tissue
RT localization.";
RL J. Biol. Chem. 276:13340-13347(2001).
CC -I- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN LIVER,
CC SMALL INTESTINE, COLON, SALIVARY GLANDS, STOMACH AND APPENDIX.
CC -I- SIMILARITY: CONTAINS 65 PLECTIN REPEATS.
CC -I- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC DR EMBL; AB051895; BAB40803.1;
DR Genew; HGNC:15577; EPPK1.
DR InterPro; IPR001101; Plectin_repeat.
DR Pfam; PF00681; Plectin; 36.
DR SMART; SM00250; PLEC; 65.
KW Coiled coil; Repeat; Structural protein; Cytoskeleton.
FT REPEAT 9 46 PLECTIN 1.
FT REPEAT 47 84 PLECTIN 2.
FT REPEAT 85 122 PLECTIN 3.
FT REPEAT 123 160 PLECTIN 4.
FT REPEAT 166 200 PLECTIN 5.
FT REPEAT 253 290 PLECTIN 6.
FT REPEAT 291 328 PLECTIN 7.
FT REPEAT 330 366 PLECTIN 8.
FT REPEAT 367 404 PLECTIN 9.
FT REPEAT 521 558 PLECTIN 10.
FT REPEAT 578 615 PLECTIN 11.
FT REPEAT 616 653 PLECTIN 12.
FT REPEAT 654 691 PLECTIN 13.
FT REPEAT 692 729 PLECTIN 14.
FT REPEAT 733 767 PLECTIN 15.
FT REPEAT 840 878 PLECTIN 16.
FT REPEAT 898 935 PLECTIN 17.
FT REPEAT 936 973 PLECTIN 18.
FT REPEAT 974 1011 PLECTIN 19.
FT REPEAT 1012 1049 PLECTIN 20.
FT REPEAT 1214 1251 PLECTIN 21.
FT REPEAT 1252 1289 PLECTIN 22.
FT REPEAT 1290 1327 PLECTIN 23.
FT REPEAT 1328 1365 PLECTIN 24.
FT REPEAT 1366 1403 PLECTIN 25.
FT REPEAT 1539 1576 PLECTIN 26.
FT REPEAT 1577 1614 PLECTIN 27.
FT REPEAT 1615 1652 PLECTIN 28.
FT REPEAT 1653 1690 PLECTIN 29.
FT REPEAT 1694 1728 PLECTIN 30.
FT REPEAT 1865 1902 PLECTIN 31.
FT REPEAT 1903 1940 PLECTIN 32.
FT REPEAT 1941 1978 PLECTIN 33.
FT REPEAT 1979 2016 PLECTIN 34.
FT REPEAT 2017 2054 PLECTIN 35.
FT REPEAT 2192 2234 PLECTIN 36.
FT REPEAT 2235 2272 PLECTIN 37.
FT REPEAT 2273 2310 PLECTIN 38.
FT REPEAT 2311 2348 PLECTIN 39.
FT REPEAT 2352 2386 PLECTIN 40.
FT REPEAT 2726 2768 PLECTIN 41.
FT REPEAT 2769 2806 PLECTIN 42.
FT REPEAT 2807 2844 PLECTIN 43.
FT REPEAT 2845 2882 PLECTIN 44.
FT REPEAT 2886 2920 PLECTIN 45.
FT REPEAT 3260 3302 PLECTIN 46.
FT REPEAT 3303 3340 PLECTIN 47.
FT REPEAT 3341 3378 PLECTIN 48.
FT REPEAT 3379 3416 PLECTIN 49.
FT REPEAT 3420 3454 PLECTIN 50.
FT REPEAT 3794 3836 PLECTIN 51.
FT REPEAT 3837 3874 PLECTIN 52.
FT REPEAT 3875 3912 PLECTIN 53.
FT REPEAT 3913 3950 PLECTIN 54.
FT REPEAT 3954 3998 PLECTIN 55.
FT REPEAT 4328 4370 PLECTIN 56.
FT REPEAT 4371 4408 PLECTIN 57.
FT REPEAT 4409 4446 PLECTIN 58.
FT REPEAT 4447 4484 PLECTIN 59.
FT REPEAT 4488 4522 PLECTIN 60.
FT REPEAT 4862 4904 PLECTIN 61.
FT REPEAT 4905 4942 PLECTIN 62.
FT REPEAT 4943 4980 PLECTIN 63.
FT REPEAT 4981 5018 PLECTIN 64.
FT REPEAT 5022 5056 PLECTIN 65.
FT DOMAIN 1935 COILED COIL (POTENTIAL).
```

FT DOMAIN 2593 2624 COILED COIL (POTENTIAL) .
FT DOMAIN 3127 3158 COILED COIL (POTENTIAL) .
FT DOMAIN 3661 3692 COILED COIL (POTENTIAL) .
FT DOMAIN 4195 4226 COILED COIL (POTENTIAL) .
FT DOMAIN 4729 4760 COILED COIL (POTENTIAL) .
SQ SEQUENCE 5065 AA; 553078 MW; 378A472DEDA19489 CRC64;

Query Match 29.2%; Score 226; DB 1; Length 5065;
Best Local Similarity 1.9%; Pred. No. 11;
Matches 87; Conservative 25; Mismatches 36; Indels 4322; Gaps 26;

QY 1 AAGGI----- 5
| | |
Db 131 ATGGLVDPAGVLAPEPACHQGLDRETHKLSLEPGTGLRFLNPTLRLTYHQLL 190
QY 6 ----- 5
| | |
Db 191 ERCVRAPSGGLALLPLKITFRSMGAVSAAELLEVGILDEQAVOGLREGRLAAVDVSARA 250
QY 6 ----- 5
| | |
Db 251 EVRRYLEGTGVSAGVVLPEGHKKSFFQAAATEHLLPMGTALPLLEAAQATHLVDPITGQ 310
QY 6 ----- 13
| | |
Db 311 RLWDEAVRAGLVSPHELHEQLLVAEQAVTGHHDPFSGSIPLFQAMKGLVDRPLALRL 370
QY 14 ----- 13
| | |
Db 371 DAQLATGLVCPARLRPLLEAALRCGLDEDTORQLSQAGSFSDGTHGGLRYEQLLALC 430
QY 14 ----- 13
| | |
Db 431 VTDPETGLAFPLSGPRGEPQPPFIKYSTRQALSTATATSVSKFRGRPVSLWELLF 490
QY 14 ----- 13
| | |
Db 491 SEAISSQRAMLAOOYQEGTSLSEKLAELSATLEQAAATARVTFSGLRDVTVPGLLKA 550
QY 14 ----- 13
| | |
Db 551 EIIDQDLYERLEHGOATAKDVGLASAOQRYLQGTGCIAGLLPGSOERLSIYEARCKGLL 610
QY 14 ----- 18
| | |
Db 611 RPTALILLEAAQATGFIIDPKANKGHSVEEALRAAVIGPDVFAKLLSAERAVTGYTDPY 670
QY 19 ----- 23
| | |
Db 671 TGOQISLFOAMQKGLIVREHGIRLLEAQIATGGVIDPVHSHRVPVDVAYRRGYFDQMLNL 730
QY 24 ----- 47
| | |
Db 731 ILLOPSDDTKGFFDPNTHENTYLLQLLERCVDPEPETGLYLLPLSSOSPLVDSATQAFQ 790
QY 48 ----- 47
| | |
Db 791 NLLSVKYRGFOQKVSAMELINSEYFSEGRRLRLRRYRQREVTLGQVAKLLEAETQRQ 850
QY 48 ----- 47
| | |
Db 851 ADIMLPALRSRVTHQLEAGIIDQQLDQVLAGTISPEALLMDGVRRLYLCGLGAVGV 910
QY 48 ----- 47
| | |
Db 911 RLLPSGQRLSYQAMRQKLGPRVALALLEAAQATGTIMDPHSPESLSVDEAVRRGVGP 970
QY 48 ----- 51
| | |
Db 971 ELYGLRAEGAIAGRDPFSGKQSVFQAMKGLIPWEQAARLLEAQVATGGIIDPTSH 1030
QY 52 ----- 57
| | |

Db 1031 HHLMPVAIQRGYVDOEMETALSSSETPTPDGQGRTSYAQLLBECPRDSTGLHLPL 1090
QY 58 V----- 58
| | |
Db 1091 PESAPALPTEEQVRSLOAVPGAKDGTSLWDLSSCHFTTEORRGILEDVQEGRTTVPOL 1150
QY 59 ----- 58
| | |
Db 1151 LASVQRWVOETKLLAQAQVVMVPGRGEVPAVWLLDAGIITQETLEALAOGTSPAQVAEQ 1210
QY 59 ----- 66
| | |
Db 1211 PAVKACLWGTGCVAGVLLQPSGAKASIAQAVRDGLPTGLQRLLEAQVSGFLVDPLNN 1270
QY 67 ----- 66
| | |
Db 1271 ORLSVEDAVKVGVLVGRSELSEQLQAERAAAGYDPYVSRAASLSLQWAMEKGLVPONEGLPL 1330
QY 67 ----- 66
| | |
Db 1331 LQVLATGGVDPVHGHLPOAAACRLGLLDTQTQSOVLTAVDKDNKFFPDPDSARDQVYQ 1390
QY 67 ----- 66
| | |
Db 1391 QLRERCVCSETGLLLPLPSTVLEVDHTAVALKAMKVPVSTGRFKGCSVSLMDLLS 1450
QY 67 ----- 66
| | |
Db 1451 EYVGADKRRELVALCRSGRAAALRQVVSATLVAEAEERQLOATFRGLRKQVSARDLPR 1510
QY 67 ----- 66
| | |
Db 1511 AQLSRKTLDELDSQTTTTVYKEVEMDSVKRSLEGNFIAGVLIQGTQERMSIPEALRHHI 1570
QY 67 ----- 66
| | |
Db 1571 LRPGTALVLEAAQATGFIIDPAENRKLTVBEAFKAGMFGKETYVKLLSAERAVTGYTDP 1630
QY 67 ----- 66
| | |
Db 1631 YTGQISLFOAMQKGLIVREHGIRLLEAQIATGGIIDPVHSHRVPVDVAYRCGYFDEMN 1690
QY 67 ----- 73
| | |
Db 1691 RILADPSDDTKGFFDPNTHENTYLLQLLERCVDPEPETGLYLLQIKGENYVYINEATRH 1750
QY 74 ----- 73
| | |
Db 1751 VLQSR TAKMRVGRFADQVVSFMDLLSSPYFTEDRKRELIQYGAOSGGLEKLEITTTI 1810
QY 74 ----- 73
| | |
Db 1811 EETETQNOGIKVAATRGVTAADLFNSRVIDOKTLTLRVGRTGGQALSTLECVKPYLEG 1870
QY 74 ----- 81
| | |
Db 1871 SDCIAGVTVPSTREVMSLHESRKELIPAAAFATWLLLEAAQATGFLDPCTRKLSVDEAV 1930
QY 82 ----- 81
| | |
Db 1931 DVGLVNEELRERLLKAERATGYRDPATGDTIPLFOAMQKGLIEKAEALRLEVOVATGG 1990
QY 82 ----- 81
| | |
Db 1991 VIDPQHHRLLPLETAYRRGCLHKDIYALISDQKMRKRFVDPNTOEKVSYRELOQCRCPQ 2050
QY 82 ----- 83
| | |
Db 2051 EDTGWLPFNKAARDSEHIDDETRALEAEQVEITVGRFRGQKPTLWALLNSEYVTEBK 2110
QY 84 ----- 87
| | |
Db 2111 KLQLVMYRTHTRRALQTVAAQILLEIEKQETSNKHLWFOGIRROITASELLSSAIITEE 2170
| | |

```
Qy 88 ----- 87
Db 2171 MLQDLETGSTTQELMEDDRVKRYLEGTSIAGVLVPAKDQPGRQEKMSIYQAMWKGVL 2230
Qy 88 ----- 87
Db 2231 PGTALVLEAQAATGFVIDPVRNRLSVVEEPVAGVVGSEIOEKLISAERAVTGYTDPYT 2290
Qy 88 ----- 87
Db 2291 GQOISLFOAMQKDLIVREHGIRLLEAQIATGGVIDPVHSHRVPVDVAYRRGYFDEEMNRV 2350
Qy 88 --NPEDDT----- 93
      :|||
      LADPSDTKGFPDNTHENTYVQLLRRCVPDPDTGLYMLQLAGRSAVHQLSEELRCAL 2410
Qy 94 ----- 93
Db 2411 RDARVTPGSGALQGQSVSWELLFYREVSEDRRQDLLSRYRAGTLTVEELGATLTSLLAQ 2470
Qy 94 ----- 93
Db 2471 AQAQARAEAGSPRDPREALRAATMEVKVGRGRVAVPVDVLAGSVYVRAAREELLA 2530
Qy 94 -----DPG----- 96
      :||
      EFGSGTLDLPALTRRLTAIEEAEAPGARPOLQDARRGPREPGPAGRGDGSRSOREG 2590
Qy 97 ----- 96
Db 2591 QGEGETOEAATAARRQEQTLRDATMEVQGFQGRPVSVMDVLFSSYLSEARRDELL 2650
Qy 97 --HA-----DLVLYITR----- 106
      |||
      AQAHAAGALGPLAVLTVRIETEERLSKVSFRGLRQVSASELHTSGILGPETLRDLA 2710
Qy 107 ----- 106
Db 2711 QGKTLOEVTEMDSVKRYLEGTSIAGVLVPAKDQPGRQEKMSIYQAMWKGVL 2770
Qy 107 ----- 106
Db 2771 LLEAQAATGFVIDPVRNRLSVVEAAGVVGSEIOEKLISAERAVTGYTDPVYTGQOISL 2830
Qy 107 ----- 106
Db 2831 FOAMQKDLIVREHGIRLLEAQIATGGVIDPVHSHRVPVDVAYRRGYFDEEMNRVLADPSD 2890
Qy 107 ---FD----- 108
      ||
      DTKGFPDNTHENTYVQLLRRCVPDPDTGLYMLQLAGRSAVHQLSEELRCALRDARVT 2950
Qy 109 ----- 108
Db 2951 PGSGALQGQSVSWELLFYREVSEDRRQDLLSRYRAGTLTVEELGATLTSLLAQAQAAR 3010
Qy 109 ----- 108
Db 3011 AEAEAGSPRDPREALRAATMEVKVGRGRVAVPVDVLAGSVYVSGAAREELLAEFSGT 3070
Qy 109 LELP-----DGN----- 115
      :||
      LDLPALTRRLTAIEEAEAPGARPOLQDARRGPREPGPAGRGDGSRSOREGQEGT 3130
Qy 116 ----- 115
Db 3131 QEAAAAAAAAARROEQTLRDATMEVQGFQGRPVSVMDVLFSSYLSEARRDELLAQAHAAG 3190
Qy 116 -----ROV----- 118
      |||
      ALGPLAVLTVRIETEERLSKVSFRGLRQVSASELHTSGILGPETLRDLAOGTKTL 3250
```

```
Qy 119 ----- 118
Db 3251 QEVTEMDSVKRYLEGTSIAGVLVPAKDQPGRQEKMSIYQAMWKGVL 3310
Qy 119 -----RGVT----- 122
      ||
      ATGFVIDPVRNRLSVVEAAGVVGSEIOEKLISAERAVTGYTDPYTGQOISLFOAMQK 3370
Qy 123 ----- 122
Db 3371 DLIVREHGIRLLEAQIATGGVIDPVHSHRVPVDVAYRRGYFDEEMNRVLADPSDDTKGFF 3430
Qy 123 ----- 122
Db 3431 DPNTHEINTYVQLLRRCVPDPDTGLYMLQLAGRSAVHQLSEELRCALRDARVTPGSGAL 3490
Qy 123 ----- 122
Db 3491 QGQSVSWELLFYREVSEDRRQDLLSRYRAGTLTVEELGATLTSLLAQAQAARAEAG 3550
Qy 123 -----QLGGACSTW----- 132
      :||
      SPRDPREALRAATMEVKVGRGRVAVPVDVLAGSVYVSGAAREELLAEFSGTLDLPAL 3610
Qy 133 ----- 132
Db 3611 TRRLTAIEEAEAPGARPOLQDARRGPREPGPAGRGDGSRSOREGQEGETOEAATA 3670
Qy 133 ----- 132
Db 3671 AAAARRQEQTLRDATMEVQGFQGRPVSVMDVLFSSYLSEARRDELLAQAHAAGALGPLD 3730
Qy 133 ----- 132
Db 3731 LVAVLTVRIETEERLSKVSFRGLRQVSASELHTSGILGPETLRDLAOGTKTLQEVTEM 3790
Qy 133 -----SCL----- 135
      |||
      DSVKRYLEGTSIAGVLVPAKDQPGRQEKMSIYQAMWKGVL 3850
Qy 136 ----- 135
Db 3851 DPVRNRLSVVEAAGVVGSEIOEKLISAERAVTGYTDPYTGQOISLFOAMQKDLIVRE 3910
Qy 136 ----- 135
Db 3911 HGIRLLEAQIATGGVIDPVHSHRVPVDVAYRRGYFDEEMNRVLADPSDDTKGFPDPNTH 3970
Qy 136 -----ITED----- 139
      :||
      WVELFYREVSEDRRQDLLSRYRAGTLTVEELGATLTSLLAQAQAARAEAGSPRDP 4090
Qy 140 ----- 139
Db 4091 REALRAATMEVKVGRGRVAVPVDVLAGSVYVSGAAREELLAEFSGTLDLPALTRRLTA 4150
Qy 140 ----- 139
Db 4151 IIEEAEAPGARPOLQDARRGPREPGPAGRGDGSRSOREGQEGETOEAATAAARR 4210
Qy 140 ----- 139
Db 4211 QEQTLRDATMEVQGFQGRPVSVMDVLFSSYLSEARRDELLAQAHAAGALGPLDVAVULT 4270
Qy 140 ----- 139
Db 4271 RVIEETEERLSKVSFRGLRQVSASELHTSGILGPETLRDLAOGTKTLQEVTEMDSVKRY 4330
Qy 140 -----TGF----- 142
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF007865; AAC06346.1; -;
 HSP; P14687; IAMU.
 InterPro: IPR000873; AMP-bind.
 InterPro: IPR001242; Condensatn.
 InterPro: IPR003880; Ppantne_attach.
 Pfam: PF00501; AMP-binding; 5.
 Pfam: PF00550; pp-binding; 5.
 Pfam: PF00668; Condensation; 5.
 PRINTS; PR00154; AMPBINDING.
 PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
 PROSITE; PS00455; ACP_DOMAIN; 5.
 PROSITE; PS50075; ACP_DOMAIN; 5.
 Ligase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
 Multifunctional enzyme; Repeat.
 FT REPEAT 39 612 DOMAIN 1 (ISOLEUCINE-ACTIVATING).
 FT REPEAT 1109 1648 DOMAIN 2 (CYSTEINE-ACTIVATING).
 FT REPEAT 2124 2689 DOMAIN 3 (LEUCINE-ACTIVATING).
 FT REPEAT 3164 3732 DOMAIN 4 (GLUTAMINE-ACTIVATING).
 FT REPEAT 4668 5249 DOMAIN 5 (ISOLEUCINE-ACTIVATING).
 FT DOMAIN 621 1037 CYCLIZATION (POTENTIAL).
 FT DOMAIN 544 611 ACYL CARRIER (ACP) 1.
 FT DOMAIN 1585 1652 ACYL CARRIER (ACP) 2.
 FT DOMAIN 2621 2688 ACYL CARRIER (ACP) 3.
 FT DOMAIN 3664 3730 ACYL CARRIER (ACP) 4.
 FT DOMAIN 5171 5238 ACYL CARRIER (ACP) 5.
 FT BINDING 574 574 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 1615 1615 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 2651 2651 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 3694 3694 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 5201 5201 PHOSPHOPANTETHEINE (BY SIMILARITY).
 SQ SEQUENCE 5255 AA; 598254 MW; 906E8DD68450F85B CRC64;

Query March 29.04; Score 224; DB 1; Length 5255;
 Best Local Similarity 1.64; Pred. No. 15;
 Matches 77; Conservative 34; Mismatches 36; Indels 4568; Gaps 24;

QY 2 AGG----- 4
 DB 110 AGGALPIDETPKDRIAFMLSDTKAAVLLTQGAADGIDCEADIVQLDREASDGFSEK 169
 QY 5 ----- 4
 DB 170 LSSVNSGDTAYIIYTSGSTGTPKGVITHYSVIRVVONTNVIDITEDNVILQLSNYSF 229
 QY 5 ----- 4
 DB 230 GSVDFDIFGALLNGASLYMIEKALLNLRGAINBEKVSVMPTTALFNMIADIVDCL 289
 QY 5 ----- 8
 DB 290 SNLRKILFGGERASIPHRVKVNLHVGKDLIHVVGTESTVYATYTYFINEIDAEETIP 349
 QY 9 ----- 8
 DB 350 GSPLANTSVLIMDEAGKLVPIGVGELCIAGDGLSKYLNREELTAETKPIPHPIPERL 409
 QY 9 ----- 8
 DB 410 YKTGDLAKWPDGNIETIFGRIDHQVIRGFRIELGEIESLEMHEDINETIVTVREDES 469
 QY 9 ----- 8
 DB 470 RPYICAYITANREISLDELKGLCEKLPYMIYPAFVKLKLPLTKNGKVDKALPEPDR 529
 QY 9 ----- 21

4331 LECTSCIAGVLVPAKDPQGRQEKMSIYQAMKGVLRPGTALVLLLEAQAATGFDVDPVRL 4390
 QY 143 ----- 142
 DB 4391 RLSVEAAGVVGGEIQEKLKLSAERAVTYGTYDPTVGQOISLFQAMQKDLIVREHGIRLL 4450
 QY 143 ----- 142
 DB 4451 EAQIATGGVIDPVHSHRVPDVVAYRRGYFDEEMNRVLADPSDDTKGFPDPNTHENLTYYQ 4510
 QY 143 ----- 142
 DB 4511 LLRRCPDPTGLYMLQLAGRSVAHQSELRALRDARVTPGSGALQGOSVSWELLF 4570
 QY 143 -----DLGVITI 148
 DB 4571 YREVSEDRDRLLSRYRAGTLTVEELGATL 4600

RESULT 4
 BACA_BACLI STANDARD; PRT; 5255 AA.
 AC 068006;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Bacitracin synthetase 1 (Bac) [Includes: ATP-dependent isoleucine
 DE adenylation (ileA) (isoleucine activase); ATP-dependent cysteine
 DE adenylation (CysA) (cysteine activase); ATP-dependent leucine adenylation
 DE (LeuA) (leucine activase); ATP-dependent glutamate adenylation (GluA)
 DE (Glutamate activase); ATP-dependent isoleucine adenylation (IleA)
 DE (isoleucine activase); Glutamate racemase (EC 5.1.1.3)].
 GN BACA.
 OS Bacillus licheniformis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OC NCBI_TaxID=1402;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 10716;
 RX MEDLINE=98089193; PubMed=9427858;
 RA Konz D., Kiens A., Schoergendorfer K., Marthiel M.A.;
 RT "The bacitracin biosynthesis operon of Bacillus licheniformis ATCC
 RT 10716: molecular characterization of three multi-modular peptide
 RT synthetases.";
 RL Chem. Biol. 4:927-937(1997).
 CC -1- FUNCTION: ACTIVATES FIVE AMINO ACIDS, INCORPORATES TWO D-AMINO
 CC ACIDS, RELEASES AND CYCLIZES THE MATURE BACITRACIN.
 CC -1- CATALYTIC ACTIVITY: L-glutamate = D-glutamate.
 CC -1- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
 CC (POTENTIAL).
 CC -1- PATHWAY: Cyclic peptide antibiotic bacitracin biosynthesis.
 CC -1- SUBUNIT: LARGE MULTIZENME COMPLEX OF BAI, BAI2 AND BAI3.
 CC -1- DOMAIN: CONSISTS OF FIVE MODULES AND ONE EPIMERIZATION DOMAIN IN
 CC THE FOURTH MODULE. EACH MODULE INCORPORATES ONE AMINO ACID INTO
 CC THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
 CC RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOLATION, CONDENSATION
 CC (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND
 CC N METHYLATION (OPTIONAL).
 CC -1- MISCELLANEOUS: BACITRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
 CC DODECAPETIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
 CC ABUNDANT IS BACITRACIN A, A BRANCHED CYCLIC DODECAPETIDE. IT
 CC CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-
 CC GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION
 CC PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-
 CC HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
 CC TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
 CC CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
 CC PHE-9, AND ASP-11).
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.

Db 530 TAGAENEYAPRNETEEKLAAMVQDVLHVEKAGIHDHFAQMGHSLUAMELIAKIKKXN 589
QY 22 -----HOED----- 25
Db 590 VEIPLHQLFKLATIKELSAFIENHOEDGDTLVTRAADPENIHEIFPLTGTQLAYLVGR 649
QY 26 -----TER-- 28
Db 650 DETFEIGGVATNLVFEADVDLNRFLTQKLI DRHPILRTIVFENGOKILEATQRT 709
QY 29 ----- 28
Db 710 IETQDLRGFTBEEINVRILEOREKMTSKIIDPSWPLFELKTFMLPGEKKYFFLNVDPLI 769
QY 29 -----YVLTNLNI----- 36
Db 770 CDDSSMKRLIREFKQLYENPGLQPSLEYSFRDYVLASINFKQTSRYQKXQDQWLDKLDH 829
QY 37 ----- 36
Db 830 FPSAPELPLKSDPAHVAKPSFKKSTFLDGHWTWNLKKKARHHHLTPTSVLCAAYAYILA 889
QY 37 ----- 36
Db 890 YMSRONHFAINLTVFNRI PFHPDVKNMGDTLSMLLDIHAEBNMSSFWRFALNVQDTLL 949
QY 37 -----GAELLR----- 42
Db 950 EALEHRRYDGVDIRNIAKNGKNKKAAMPVFTSVLSENPDSDSLVDFDNHFFSTR 1009
QY 43 ----- 42
Db 1010 TSQVYIDNQVEINGLYITWDYVEQI FEHEVIESMFDQYIAVIQKAVGEDVSTIQMNE 1069
QY 43 ----- 42
Db 1070 KSRQIMISAYNDTQSDPAKPLHELFTGQVKGHPDRMAKHHDEVMTYOELDEKSNQVAF 1129
QY 43 -----DP----- 44
Db 1130 LIGKGVKGDYIGVIGKRSGLTGNLLAVLKTGAAYIPLDPDYPERKAYIOSKSNCKFF 1189
QY 45 ----- 44
Db 1190 ISHDVYDKEHIERFSKAPVDRKVDLDDMAYVIFTSGTGKPGVQITHVPQRNTILDINE 1249
QY 45 -----SLGAQFRV----- 52
Db 1250 KFNVTQDNIMGSSLCFDLSVYDVFCAISSGASLVIIDQDRDVFSLKETAEKERITIW 1309
QY 53 -----HLVKMWIL----- 60
Db 1310 SVPAIMGMTADVDPDNELNHLRLIILSGDWIPLQLPATIKKTFKNAEVISLGGATEGSI 1369
QY 61 ----- 60
Db 1370 WSIYPIQKVEEDKSI PYGKPLANQKIYVLNQNKLCPVGVEGELYIGGAVASGYIHD 1429
QY 61 ----- 60
Db 1430 OEKTEHSFIOHOBELGYIYKTGDYVLKEDGYVEFLGRKDSQVKIRGYRVMGEIENTLVS 1489
QY 61 -----TEPEGAPNITA----- 71
Db 1490 HOEITKASVIDYTPDGIKNLYAFVVAENAI SLDVKEFLQKTLPDYMI PAKFVQIEEIP 1549
QY 72 ----- 71
Db 1550 LTVNGKVDKRTLHLAEQHTADEGQRGRLPENETQAMLEIWKDIFGLDSINLDVSY 1609
QY 72 ----- 71

Db 1610 EIGDSLKASIIITEINKRMVEMPISEIFKNDTIIALDHLKKNRESDMEHPQKAREK 1669
QY 72 ----- 71
Db 1670 EYPTSPAQRMYSMLENERGAYHIPALLVEGRINAMQLENALKTFLOREILRTGF 1729
QY 72 -----NLTS----- 75
Db 1730 EIONNELIQIYENVDFREYECLDASITDQHALMEITSRYCKESIKFPDLSRPLMRAK 1789
QY 76 ----- 75
Db 1790 LIKIDDIRHILVINPHHII SDGVSQILMNEILELSNVPLPEVNVQYKDYVEMHHTFNQ 1849
QY 76 ----- 75
Db 1850 SAAMKKQAYWLDVYRIDPSKLDPPDYKRHHIDTFEGSSVPLEMERELSDHIRKLAKHN 1909
QY 76 ----- 75
Db 1910 GTTLYTVMLSAYYYVLLNKYTNQTDIVGTAAGRHLHPLQDQVGVFVNTLALRNEVDTSY 1969
QY 76 ----- 75
Db 1970 SFKEFLOQTERTIAAFDENSEYPPDDLLIRKLVGRESNRNPLFDTMFVLEDARMFTKQK 2029
QY 76 ----- 75
Db 2030 DVKSPILFELDNAKFDIMFNVLDFEQKIVLNIESTLSLKDETIQKIAEDYFRILEEVS 2089
QY 76 ----- 75
Db 2090 ENLDVALHQIDMISROEKRTLLESFNHTKTAYPKGKAHQHFEEQAKRIPDHTAVVFEDQ 2149
QY 76 ----- 75
Db 2150 KLTYQLNEKANQVALLREKGVKPDTLVGIMMERSSDMAAILGLVXAGGAYLPIDPEY 2209
QY 76 ----- 75
Db 2210 PPERMYAFDSEVKVVIISDVPLAELTAESI ELIHMDDERIAGDRSDIDNVNOSGDLA 2269
QY 76 -----SLSVCGHSG----- 85
Db 2270 VVIYTGSGTKPGVMI EHQSLINLCSMHQSCFEVQNDNSSIYASISFDPFVWELFPYI 2329
QY 86 ----- 85
Db 2330 TAGATVHVNLNOETRLDVEKLNRYFHDHHITISFLPTPVEQFTALDNLHSLRLLTGGDKL 2389
QY 86 ----- 85
Db 2390 NVFKEKSYQIVNNYGTENTVATSPIDKSHQNIPIGKPIDNVKVYILNKDQLCPLQA 2449
QY 86 ----- 85
Db 2450 SGELCIAGEGLARGYVNRPELTREKFI GNPFPVGPGRMYRTGLAKMLPDGNIQFLGRVDQ 2509
QY 86 ----- 85
Db 2510 QVKIRGYRIEPEIEIENRLLKYEIEEAAVIAREDGDHDPYLCAYVTVKKEVEPEKIRAF 2569
QY 86 -----TIN-----PED----- 91
Db 2570 KKSPLDYMIPOYFVQLDGLPLTVNGKVDKSLPVSPERSVTMDRRYEAPRDOMEEKLVSIW 2629
QY 92 ----- 91
Db 2630 EEALGINKIGINSHPFAGGHSKAAALVSTIHKELNVKLPLURQIFETPTIKGLRDISVR 2689
QY 92 ----- 91
Db 2690 RRKCFYIDRTEEKPYRSLSSAQKELYILSQTGSHVAYNMPFAMTLEGDFDIRRENTLUK 2749

```
QY 92 ----- 91
Db 2750 NMVKRHSFRTFWMIDGVNQIEKIDFOVAYSDIGKSAEBKISFIRPFHLEKAPL 2809
QY 92 ----- 91
Db 2810 LRAEVVKLNEREHLMLFDMHHIISDGVSDFIFIQELGALYEGKSLKPFHIQYKDYAEWEN 2869
QY 92 ----- 91
Db 2870 SHARSEELKQEEYWLTKYKGDIPVLDLPIDHKRPLTKSSEGDVTAAIESETPRKLQHM 2929
QY 92 ----- 91
Db 2930 AKENGVTMYMLLAGYTALLSKYTQEDIIIVGTPAAGRHNHEDIOHLIGMFVNTLAIRNP 2989
QY 92 ----- 91
Db 2990 EGKKTFRDYLOEVKENTLOAYENODYPFEELVEKVNIKRDMARNPLFDTMLVYHNTDVKP 3049
QY 92 ----- 91
Db 3050 FEAGLSRLVEIKRGISKEDITVTASEAADGLRLEVEYSTTLFNKERMERLSEHLISLL 3109
QY 92 ----- 91
Db 3110 EQAADHPDIAINOIDVLTGERHRVLYDFNRTDGVFCCKEMTIPELFEKQAEKTPDHPAVA 3169
QY 92 ----- 91
Db 3170 FGDETI SYRELNERANS LAFTLROKGVGPVGIAGILTERSIEMIVGIMILKAGGAYLPI 3229
QY 92 ----- 96
Db 3230 DPAYQERISYIVKSDSVLCAAGVDVPGEAYTGDIIRIDQTQONDHVENLKHDIKPOH 3289
QY 97 ----- 96
Db 3290 LAYVIYTGSTGKPGVMIEHHSVNNLVHGLNERIYQHLDHNLVALVAPYIFDASVKQI 3349
QY 97 ----- 110
Db 3350 PAALLFGHTLCIVPRETAWDAMSLIEYYSKNNINVSMDTPAHLNMLAYVDKTELEFDVKE 3409
QY 111 ----- 110
Db 3410 LIVGDALTPDVIGGLFHKFPNLSNITNVYGPTECCVDAASHQIESGKVPQTPSIPIGR 3469
QY 111 ----- 110
Db 3470 PLLNTSIYVDKELRPLPVGIAGELCIAGEGARGVNNRPELTAEKFVDHPFEPFKMYK 3529
QY 111 ----- 114
Db 3530 TGDMLMLPDCQIEFLGRADHQVKIRYRIELGEVEQQLTHEKIKEAAVIAGKDQNGNS 3589
QY 115 ----- 114
Db 3590 YLCAYIASDKELPAADVQFLEREMPDYMIPSYFVKLORLPRTSGKVDRSALPEADGNV 3649
QY 115 ----- 114
Db 3650 NVMEGTGYPNREIERKLQVWREILGAEDIGISHHFFAAGGDSIKALQIVSRLAKMNL 3709
QY 115 ----- 124
Db 3710 KLEMKALFANPKIKDLSRFITEETHRKHNPVTGETELLPIQKRYFANNKEELDHFNQS 3769
QY 125 ----- 124
Db 3770 FMLFRKGDYDENIVPTAFNKILEQHDALRMIYEKDDGDIQVNRGYRENLFOLDVYVRG 3829
```

```
QY 125 ----- 124
Db 3830 FDSQEEKVPELATGIQKSSIRKGLVHLGIIPRADEGDHLLIAIHHLVWDGSMRILFED 3889
QY 125 ----- 124
Db 3890 FETLYQALKGEPLDIGYKTDSYOEFARQLKQYAOQRLLKEREYWKALEADVFPFPAE 3949
QY 125 ----- 124
Db 3950 KLERDFFHSATLSIRIGDPDVTAKLLRNAPKAYNTEINDILLTALIAAVRDIITGENKLV 4009
QY 125 ----- 124
Db 4010 MMEGHGREDLDGVDITRTIGWFTTVYPFIDLGEKEISONIKMWKEALRKIPNKGIGY 4069
QY 125 ----- 124
Db 4070 GVLKYMTEELQKIQTOAPLSFNFGEMNNDMNRKVSQSPSPGESIGGKIVRHCIAEMN 4129
QY 125 ----- 131
Db 4130 AISLNGELTIYTFNQDQYOTSTIEOLNOSFKENLEKIVDHCVDKESDMTTPSDYGVSL 4189
QY 132 ----- 131
Db 4190 GLSELELIKDYSAFOIEKIYPLANMOKGMLPHNAMDQTSQAYFQQIVIKLGRVHPDIL 4249
QY 132 ----- 131
Db 4250 EESFHEIVKHEILRASFEYEITAEPRQIIARDKTPFTSIDLTGENRTRQHRPIETYLK 4309
QY 132 ----- 133
Db 4310 EDOEKGFDLSSEALMRVCLIKMSDESRYRLIWSHHILLDGCGLIVLSELPSLYGKIMKG 4369
QY 134 ----- 134
Db 4370 ESRLKEPKPYGDIYKWLKQDOEAVAYWKDYKGVESRSELPAPNRGATSEECCKEK 4429
QY 135 ----- 134
Db 4430 VISFSKELTTKTRIAKOHVHTINTVLQIGWIMILAKYKNTDEVFQTVVSGREAPVDGI 4489
QY 135 ----- 137
Db 4490 EEMVGLFIHTIPTRISFEGARSFKEVLKKTQAESIESNRYSYNNLSEIQVLSEMKRELIT 4549
QY 138 E-----DTGFDL----- 144
Db 4550 HVMAFQNYAFDEELFRSQSGETGFELEGVHGKERTNYNFNLTGVLEDEQLKLTFFENV 4609
QY 145 ----- 144
Db 4610 YDNTIETLEKHIITVAEQVAEDETQTLRDINLVKEEOHRIHLTFNDTKTGYPKDXPLH 4669
QY 145 ----- 144
Db 4670 ELFEQAMKTPDHTALVFGAORMTYRELNEKANQATARLLREKIGRGSIAAIIADRSFEM 4729
QY 145 ----- 144
Db 4730 IIGTIGILKAGTYLPIDPETPRDRIDYMLKNSGAALLVTTDSLKKPFDIKTVDLCSDEL 4789
QY 145 -----GVTI 148
Db 4790 HLLSEENLPRNRSDDTAYIVYTSGTGTPKGVVI 4824
```

```
RESULT 5
HTSI COCCA
ID HTSI COCCA
AC Q01866;
STANDARD; PRT; 5217 AA.
```

DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE HC-toxin synthetase (EC 6.3.2.-) (HTS).
GN HTS1.
OS Cochliobolus carbonum (Bipolaris zeicola).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Cochliobolus.
OX NCBI_TaxID=5017;
RN [1]
RN FUNCTION.
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 90305 / SB111;
RX MEDLINE=93100328; PubMed=1281482;
RA Scott-Craig J.S.; Panaccione D.G.; Pocard J.-A.; Walton J.D.;
RT "The cyclic peptide synthetase catalyzing HC-toxin production in the
RT filamentous fungus Cochliobolus carbonum is encoded by a
RT 15.7-kilobase open reading frame.";
RL J. Biol. Chem. 267:26044-26049 (1992).
RN [2]
RN FUNCTION.
RP STRAIN=ATCC 90305 / SB111;
RX MEDLINE=20138231; PubMed=10671527;
RA Cheng Y.-Q.; Walton J.D.;
RT "A eukaryotic alanine racemase gene involved in cyclic peptide
RT biosynthesis.";
RL J. Biol. Chem. 275:4906-4911 (2000).
CC -!- FUNCTION: Non-ribosomal peptide synthetase, able to activate
CC proline and ARO (2-amino-9,10-epoxy-8-oxodecanoic acid), and
CC epimerize L-Pro. Catalyzes the production of HC-toxin: a cyclic
CC tetrapeptide. Activates and thioesterifies L-Pro, and epimerizes
CC it to D-Pro; also uses D-Ala as a substrate but this is epimerized
CC from L-Ala by TOXG.
CC -!- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF HC-TOXIN.
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -!- SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M98024; AAA33023.1; -
DR HSP; P14687; 1AMU.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR001242; Condensatn.
DR InterPro: IPR003880; Ppantne_attach.
DR Pfam; PF00501; AMP-binding; 4.
DR Pfam; PF00550; pp-binding; 4.
DR Pfam; PF00668; Condensation; 5.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 4.
DR PROSITE; PS00455; AMP BINDING; 3.
DR PROSITE; PS00075; ACP DOMAIN; 4.
KW Ligase; Multifunctional enzyme; Phosphopantetheine; Repeat.
FT REPEAT 249 842 DOMAIN 1.
FT REPEAT 1854 2452 DOMAIN 2.
FT REPEAT 3006 3606 DOMAIN 3.
FT REPEAT 4158 4738 DOMAIN 4.
FT DOMAIN 762 840 ACYL CARRIER (ACP) 1.
FT DOMAIN 2384 2450 ACYL CARRIER (ACP) 2.
FT DOMAIN 3536 3604 ACYL CARRIER (ACP) 3.
FT DOMAIN 4667 4736 ACYL CARRIER (ACP) 4.
FT BINDING 803 803 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 2414 2414 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3568 3568 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 4700 4700 PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ SEQUENCE 5217 AA; 574637 MW; 0331D9C5400163A5 CRC64;

Query Match 28.7%; Score 222; DB 1; Length 5217;
Best Local Similarity 1.8%; Pred. No. 19;
Matches 87; Conservative 20; Mismatches 36; Indels 4702; Gaps 27;
QY 2 AGGI----- 5
DB 289 AGGVCVPIDPRYPVERIIRDIIRTNATIALVGAGKTAALFKSADTAVQITIDITKDIPHGL 348
QY 6 ----- 5
DB 349 SDTVQSNKIDDPAGFLTSGTGVKCIIVVTHSOICTAVQAYKDRFGVTSETRVLQFS 408
QY 6 ----- 5
DB 409 SYTFDISIADTFALFYGGTLCIPSEEDRMSNLQDYMVSVRPNWAVLTPTVSRFLDPGVV 468
QY 6 ----- 5
DB 469 KDFISTLIFTGEASREADTPWIEAGVNLVNYGPAENTLITTATIRKGSNIGYGVN 528
QY 6 ----- 5
DB 529 TRTWTVDSGACLVPGSIGELLIESGHLADKYLNRPORTAEAFSLDLPWIPNYEGDSVR 588
QY 6 ----- 5
DB 589 RGRPRYRTGDLVRYCDDGSLICVGRSDTQIKLAGORVELGDVEAHLQSDPTTSQAAVFP 648
QY 6 ----- 5
DB 649 RSGPLEARLIALVTGNKDGTPHNOQSLPKAPAFQCPDLVKYATSSLOORLPYSYMPVS 708
QY 6 ----- 5
DB 709 WLGDPLPMSVSGKLDRAVLQDQLESLSFSYAEILGTTGLEVDPGGAASSVASDSDLRD 768
QY 6 ----- 5
DB 769 MNDDSELLTACSRVNLNPAKISYSQSFIHAGDSITAMQVSSWMKRPFGKIGVKDLLV 828
QY 6 ----- 5
DB 829 SPSISTAASCIKSAQSGSRNFVAVRPGORIPVSPIOKLFQTAEASKSNHMYHQSFPR 888
QY 6 ----- 5
DB 889 DQPIKPTQIEDAISLMORHPMLQARFETEREGDWYQYIPIDVERRASVEVIGSLSTDDR 948
QY 6 ----- 5
DB 949 EAAMLRARQSIDLTEGLPIRCOLFNNNVDEASRLFFVVIHHAVDLVSVRIIMEELEAHL 1008
QY 6 ----- 5
DB 1009 ATDSTPDRGEAQVESVPFLAWCQVQAEAVKDIPOVDTVPLIPKIPTADFGYGLKHENV 1068
QY 6 ----- 7
DB 1069 YGNTVERKIPLGHSITEDLLYKCHDSLHTKTTIDVLLAAVLVSFRKSFDRPVPVFNVEGH 1128
QY 8 ----- 7
DB 1129 GREPGGEDAVDLSTRTVGWFTTISPVVPEVSPGDIIDVVRVKDYRWATPNNGDFYFSTK 1188
QY 8 ----- 8
DB 1189 YLTQSGIKLFEDHLPALFVNYEGRYQAMESQTVLKPESWHAGEASKQDQDGLRRFCLF 1248
QY 9 ELLVAVGPD----- 17
DB 1249 EISTAVLPDQGLHLTCSWNKNMRHQIRRLWLDLTLLPAAIGSVSSALASPOLTLSDOVE 1308

QY	18	-----	17
Db	1309	LLRLDYSSLDILKKSILSIPAVQTLDDLEGVPGSPMQDALFLSKSQSDGAYEVDFTW	1368
QY	18	-----VFQAHQED--	25
Db	1369	RVATSLQNSQPAVDIGCLVEANKDVALHAALRTVILESSLPATGILHQVVLKSHDPIV	1428
QY	26	-----TE-----	27
Db	1429	ILDVRDVTAITILDSYPPPTTEEGIALIKRPPHRLLICTTIEGSLVLIKFOVNHVLFQDMS	1488
QY	28	-----RYV-----	30
Db	1489	TDKIIQDLSKAYTCRHSNKLPHQESKLDGTYGNRPTKPLAEFIRYIRDPORKQDSIN	1548
QY	31	-----LTNLNIG--	37
Db	1549	YWKNALRGATTCSPFPLFDQITSEKAMPQSWASVPIPLCVDSKELSKTLANLIGITMST	1608
QY	38	-----	37
Db	1609	FQTVMAIVLRIYSONGQSVFGLTSGRDAPVDGIDSAVGNFIAMLVCFDFDDGQVHTA	1668
QY	38	-----	37
Db	1669	DMARKIHNASANSISHOCSLAEIQDALGLSTSTPLENTAFTYLPKRTNPKAGEPEHLL	1728
QY	38	-----AELLR 42	
Db	1729	CFEELMSDPTFDTLVEPTQESNEVSAHLDFKLKLSIQAAYSIASTVAHILSELVH	1788
QY	43	DP-----	44
Db	1789	DPYRALNTLPVSEHDTAIRSWNDHLPFPATECIHETFSRKVVEHPQREAIKCSWDGSLT	1848
QY	45	-----SLGAQFRVHLVXM-----	57
Db	1849	YAEISDLSQLSLHLSVLGKIVGTKIPICFEKSMWTIVILAVQAGGVFVLEPGHPES	1908
QY	58	-----VILTEPE-----	64
Db	1909	RLSGIIKOVQAEILLCSPATSRMGALQNIQTGMEFKIVELEPEFIRSLPLPKPNHQ	1968
QY	65	-----GAP-----	67
Db	1969	MVGLNDLYVVFVFTSGTGVKGAATHQATGIYEHAVACGMTSLGAPPRSLQFASYSF	2028
QY	68	-----NITANL-----	73
Db	2029	DASIGDIFTTLAVGGCLCPREEDRNPAGITTFINRYGVTWAGITPSSLALHLDPAVPTL	2088
QY	74	-----	73
Db	2089	KALCVAGEPLSMVTVVWSKRLNLINMYGPTTEATVACIANQVCTTTTIVSDIGRGYRATT	2148
QY	74	-----	73
Db	2149	WVQPDNHNLSVPIGAVGELIIEGSLCRGYLNDPERTAEVFIKQDQARAMEVPLFMVP	2208
QY	74	-----	73
Db	2209	TGDLVRSADGKIIFIGRKDQTVKXNGQRFELGEVEHALQLOLDFSDGPIIVDLLKRTQS	2268
QY	74	-----TSSLLS-----	79
Db	2269	GEPDLLIAFLVGRANTGTGNSDEIFATSTSSLSSEFSTVIKKLQDQARAMEVPLFMVP	2328
QY	74	-----	79
Db	2329	QAYPIEGGIPLTAAGKIDRRMLRKLCEPFNRNDLISFTSKALSTSVKDAETTTDVEDRL	2388
QY	80	-----	79

Db	2389	ARIWEKVLGVKVGRESDFSSGNGSMAAIALRAEAQRSGFTLFVADIPTNPRLADMAKL	2448
QY	80	-----	79
Db	2449	FSHQSVSPSSSLRTKVPISLQKRSSGLOTAAPVSNGPSVRRCKOKENIIDCPVAFEYE	2508
QY	80	-----VCG-----	82
Db	2509	EGPSDTOLKEASRICIGISSRSIEDVPPCTPMQOEALVALSLIPCAQASALHAFAELRPG	2568
QY	83	-----	82
Db	2569	DRNFRSAMESVYKAQPIILRSRIISGSSGVVVTSDISIPOLDVSGLDTFLEOQOVG	2628
QY	83	-----WSQ-----	85
Db	2629	FAPGAPLFLAFVYKADDCDFVISAHHAHYDGNLSNLWSQVLAALYTNNGELPPGPGSF	2688
QY	86	-----	85
Db	2689	XHFARNLNVQSKLDSDFWRKLLVKPDQESFRFPDVPVGHKPATRCTTNFHFPSMQSK	2748
QY	86	-----TIN-----	88
Db	2749	IGTANTCINAAWAILTAQYSSNKTNVFGVTLWGRDPPMIDIEMTGPTTIVTVPROVNI	2808
QY	89	-----	88
Db	2809	PESSVAEFLODLQSLAVVLPHOHLGHRHQALGPRIARQACDFSTLLVNVHGSISWSEL	2868
QY	89	-----	88
Db	2869	EAADIVPVPLRSSDLAYAPMVEVENASSDTLDIRVHSDPDCEVQLLERLMEQFGHNLQ	2928
QY	89	-----	88
Db	2929	TLCAASFDPGKRIAEMLDDTATTHTLTLFSMNSRVKSDPVAIAVHKLEETAOSOPA	2988
QY	89	-----	88
Db	2989	ESAIVAHGOLSYMQMDRCADVLAQIRKTNMISAQSPFCIHLRLRSATAVSMVLAKA	3048
QY	89	-----PED-----	91
Db	3049	GGAFMPVDISQPRSRLONLIEESCAKLVLTLPESANALATLSGLTKVIPVLSSELVQOIT	3108
QY	92	-----DTDP-----	95
Db	3109	DNTTKDEYCKSGTDPSPAYLLYTSCTSGKPKGVVMEHRAWSLGTCHLSYMGFNSCT	3168
QY	96	-----	95
Db	3169	RILQSSLMFDLSLEIWAIVYAGGCLFIPSDKERVNNLODFTFRINDINTVLTPTSIGKL	3228
QY	96	-----	95
Db	3229	LNPKDLNPISFAGFIGEPMTRSLIDAMTLPGRLLVNSYGPTEACVLVTAREISPTAPHDK	3288
QY	96	-----GHA-----	98
Db	3289	PSSNIGHALGANIWWVEPORTALVPIGAVGELCTEAPSLARCYLEANPERTEYSPSTVLD	3348
QY	99	-----DLVLYIT-----	105
Db	3349	NWQTKGTRVYRTGDLVRYASDGTDLFGRKDGQIKLRGQRIELGEIEHHIRRLMSDDPR	3408
QY	106	-----	105
Db	3409	FHEASVOLYNPATDPRDRAVDVQMRFPYLAGLLVLVLVSLMRSDMSMHVLNIANTSENL	3468
QY	106	-----	105

Db 3469 QTLVTELKSLRGVLPVPLHFVAVSLPTGSSGKLDHAFVRACLRELTAFLDGNFPK 3528
 QY 106 ----- 105
 Db 3529 VEOVLTTNESVLQWNGTVLAMPDPHSIQRGDDFFSLGGSSISAMRLVGLARSSGHKLQHE 3588
 QY 106 -----RPDL----- 109
 Db 3589 DIFMCPRLADMAQGISFVQEASVPTTPTIKFDLLDDCEVIDIHILPQDMNKELIE 3648
 QY 110 ----- 109
 Db 3649 DVYPTPLQESLMAATARHGEAYTMQISITVLAQAQLKAMDMVFRDFEVLRTIALG 3708
 QY 110 ----- 109
 Db 3709 PSQALQVVKHELSWESFSPQSKDHFYRSLGKPLARLAVITQALDTKQPSHGT 3768
 QY 110 ----- 109
 Db 3769 REARTKNSQDTVMVVGAAHHSIYDAHVLMSIMRRLYREFITGSOADGILEAETSRSSEGVVP 3828
 QY 110 -----ELPD 113
 Db 3829 FKSVEKLLRGKDNDESLLFPWKEKLRGVSSQPPASPAPRVLEHQPSTATQTLTKVSLPT 3888
 QY 114 GNRQVRGVT----- 122
 Db 3889 SSRKLGATVATVAYAAWALTIATYADPDVVFAGATLSGRETMAGSISHPESITAGTIT 3948
 QY 123 ----- 122
 Db 3949 VPLRIIDFQTVSDFLSTLQKDIVRAAYFGQMGLNSIAHIDNDCRDACGFKSIIVQV 4008
 QY 123 ----- 122
 Db 4009 PDEGENHVGRAANPFQMSLESIGHFPAPLYVEVEQSESTDVLIRMAVDPLVPEKLAHFI 4068
 QY 123 ----- 122
 Db 4069 SDTFTTMSNLNANPKAKVESIPALSEAHLAELDVTCPEWILGKAKDEKIRTESHQCLQ 4128
 QY 123 -----QLG----- 125
 Db 4129 DLVCRRAQSPNSQAIDSDMGDSISYHELDGLSSIAELHLSQLGVPEAPVCLLFEKSKWA 4188
 QY 126 ----- 125
 Db 4189 VVAMIGIIKAGGCFVPLDPSYPHERLEHIISETGSSVITVSAAYSKLCLSLSVRGIVCDG 4248
 QY 126 ----- 125
 Db 4249 SVFSSTKKPLPSTADSPSPSVRNQAAVILFTSGTGKPKGVVMEHHSVCSALIALGKR 4308
 QY 126 -----GAC----- 128
 Db 4309 MGLGQPSRVLPNSYFDMVLLDIFGLTVYGGCLCIPKESQMSNLGWSQKFKVNTMLL 4368
 QY 129 ----- 128
 Db 4369 STSVRLMQPADTPSLETCLTGEAVLQSDVDRWAPKHLIAGYGPPTCIMSVELTP 4428
 QY 129 ----- 128
 Db 4429 SSPANLIGKPVSCQAWINPLKETELAPYATGELYTQGTAVRGYLDHDDVLSKAFIVD 4488
 QY 129 ----- 128
 Db 4489 PQWLTKYTNENQWSRRAYKTGDLVFWGPOSNLYYVRKDSQVKIRQVRVLAETIEVI 4548
 QY 129 ----- 128
 Db 4549 RQHIPPVTVCVDLLSSDDQNTRIILGAVLIGDRALGGPEDLEVIGYMDLKSIIIPAL 4608

QY 129 -----SPT----- 131
 Db 4609 EASLPHHMIPEAYVPFVQLPTLGSGLKDRKTVRRVAGVPLAFSLPOASARHPNOPTVTHQ 4668
 QY 132 ----- 131
 Db 4669 KLLRLMCKILPOLDESANVKNODNFGIGGDSIAAKLVALLRQHGISLAIAEITRPTL 4728
 QY 132 ----- 131
 Db 4729 EAMSSLIDEHNFVSHAGILSDVTRNTSGVMROTTNLIAGRHSMVAKESRECDNSTLPCT 4788
 QY 132 ----- 131
 Db 4789 EYQOMFLAGTEAFTGAHSAQFIFRLPEKIDLRLOAOFDHCADWYPNURTOIHKOADTOR 4848
 QY 132 -----WSC----- 134
 Db 4849 LLHDISPIGVKVPWSCHYSDDLNLTLSHDKKPPGDLGPHLHRTIMRHRDPTESMLVWTL 4908
 QY 135 -----LITEDT----- 140
 Db 4909 NHAAYDANSLRMWLEHITEAYANPDYPSYSLGWTAFLVLTENTKEASRSFWSVSLVDK 4968
 QY 141 ----- 140
 Db 4969 PARLMFNVLNVSINPRQRLYEARIINIPKRVLSQATAATVLLAGLTLVARVCDTRDVILA 5028
 QY 141 ----- 140
 Db 5029 HLLTGRTPLAGIENCPGTTITKVPRLPILPMDQDLVLELDSVAKKITHELMRMPHEHS 5088
 QY 141 -----GFDL 144
 Db 5089 GLSAIREFIPOAEGTTSSGKPHAGSVLGRPLDLVIHPKGLDL 5133

RESULT 6
 BIR6 HUMAN
 ID BIR6 HUMAN STANDARD; PRT: 4829 AA.
 AC Q9NR09; Q9ULDI;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Baculoviral IAP repeat-containing protein 6 (Ubiquitin-conjugating
 DE BIR-domain enzyme apollon).
 GN BIR6 OR KIAA1289.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20012759; PubMed=10544019;
 RA Chen Z., Naito M., Hori S., Mashima T., Yamori T., Tsuruo T.;
 RT "A human IAP-family gene, apollon, expressed in human brain cancer
 RT cells.";
 RL Biochem. Biophys. Res. Commun. 264:847-854(1999).
 RN [2]
 RP SEQUENCE OF 3238-4829 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20039619; PubMed=10574462;
 RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
 RA Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:337-345(1999).
 CC -!- FUNCTION: MAY PROTECT CELLS FROM UNDERGOING APOPTOSIS.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN CANCER CELLS.
 CC -!- SIMILARITY: CONTAINS 1 BIR REPEAT.

Qy	66	-----	65	Qy	107	-----	106
Db	1641	VGPVHNSVPNSPAAPGFFIHPSDVIPTPKTTLFMTPLTPNEAVSVVINAELAQLF	1700	Db	2721	AHLVSDPNLIHVLVKFLSCTSPHGTNOHSPQVPTATOAMQBFTRLOVHLSSTCQPIF	2780
Qy	66	-----	72	Qy	107	-----	106
Db	1701	PGSVIDPPAYNLAHNKNSKSRMPLGSLALAISHASHFLOPPHQSIIERMHSGAR	1760	Db	2781	SEFLKLIIHLSTERGAFOTGGQPLDAQVKLLEFTLEQNEFVSVSTISAVIESVTFVLVH	2840
Qy	73	-----	82	Qy	107	-----	109
Db	1761	RFVTLDFGRFILLTDVLIPTCGDLASISIDIWTLGEEVGRRLVVATDSTHSLILHDLI	1820	Db	2841	HYITCDKYMRSRSGSSSVGARACFGLFANLIRPGDAKAVCGEMTRDQMLFLLKLVNI	2900
Qy	83	-----	83	Qy	110	--ELP--	115
Db	1821	PPVCREMKITVIGRYGSTNARAKIPLGFFYGYHTYILPWESELKLMHDPLKGESANQP	1880	Db	2901	LVQLPLSGNREYGARVSVTTNTTDSVDEBKVGKDGNGSGSTVOGSPAYVADLVLANQ	2960
Qy	84	-----	83	Qy	116	-----	115
Db	1881	EIDQHLAMVALOEDIOCRYNLACHRLETLLOSIDLPLPLSANNAAQYFLRKPDKAVEEDS	1940	Db	2961	QIMSQILSALGLCNSAMAMITGASGLHLYTKHENFHGGLDAISVGDGLFTILTTLSSKAS	3020
Qy	84	-----	91	Qy	116	-----	115
Db	1941	RVFSAYODCIQLQLNLNAHNAVORLKVAGASRKMILSETSNPEDLIQTSSTEOLRTIIR	2000	Db	3021	TVHMMLOPILTYMACYMGROGSLATCOLSEPLLWFILRVLTSDALKAFHDMGGVQLIC	3080
Qy	92	-----	91	Qy	116	-----	115
Db	2001	YLLDTLLSLHSHNGHSPAVLQSTFHAQACEELFKHLCISGTPKIRLHTGTLVLVOLCGG	2060	Db	3081	NNMVTSTRAIVNTARSMVSTIMKFLDSGPNKAVDSTLKTILASEPDNAEGIHNFAPLGT	3140
Qy	92	-----	91	Qy	116	-----	119
Db	2061	ERWGOFLSNVLOELYNEOQLIPPODRVFMLLSCIGQRSLSNGVLESLLNLDNLLSP	2120	Db	3141	ITSSSPTAOPAEVLLLOATPPHRRARSAAWSYIFLPBEAMCDLTIHLPAVLLKEIHIQPH	3200
Qy	92	-----	91	Qy	120	-----	119
Db	2121	LOPOLPMHRRTEGVLDIPMISWVMLVSRLLDYVATVEDEAAAAKPLNGNQWSFINNL	2180	Db	3201	LASLATCPSSVSVEVSADGVNMLPLSTPVVTSGLTYIKIOLVKAEVASAVCLRLHRPRDA	3260
Qy	92	-----	91	Qy	120	-----	119
Db	2181	HTOSLRSSKSSSLDRLYSRKIRKQLVHHKQQLNLLKAKOKALVEQMEKEKIOSNKSS	2240	Db	3261	STLGLSQIKLLGLTARFTTSSATVNNPFLPSPEDQVSKTSIGMURLLHCLTHISLEGMM	3320
Qy	92	-----	94	Qy	120	-----	119
Db	2241	YKLLVEQAKLQATSKHFKDLIRLRTAWSRSLNLTDEVTTAKESPEIEPLPTLAHERC	2300	Db	3321	ASAAAPTANLLOTCALLMSPYCGMHSPNIEVVLVKIGLOSTRIGUKLIDILLRNCAASG	3380
Qy	95	-----	106	Qy	120	-----	123
Db	2301	ISVVQKLVLFLLSMDFTCHADLLLFVCKVLARIANATRPITHLCEIVNEPQLERLLLLV	2360	Db	3381	SDPTDLNSPLLFRNLGSLSDSTIDILYQLGTTQDPTKDRIOALLKMWSDSARVAAMKR	3440
Qy	107	-----	106	Qy	124	-----	123
Db	2361	GTDFNRGDISWGGAWAQYSLTCMLQDILAGELLAPVAABEAGTVGDDVYGATAGSDDS	2420	Db	3441	SGRMNYMCPNSTVEYGLLMPSPSHLHVAAAILMHSYELLVEYDLPALLDQELFELLFNW	3500
Qy	107	-----	106	Qy	124	-----	123
Db	2421	LQOSSVOLLETIDEPLTHDITGAPPLSLEKXIDLELLQDLMEVDIDPLDIDLEKDLPL	2480	Db	3501	SMSLPCNMVLKXAVDSLCLSMCHVHPNYPFSLMGWNGITPPPVQCHHRLSMTDDSKKQDL	3560
Qy	107	-----	106	Qy	124	-----	123
Db	2481	AAKFKPISSTWYDWGADYGTYNPNPYIGGLGIPVAKPPANTEKNGSQTVSVVSQALD	2540	Db	3561	SSSLTDDSKNAQAPLATESHLATLASSSQSPEAIKQLLDGLPSLLVRSLSASFCSHIS	3620
Qy	107	-----	106	Qy	124	-----	128
Db	2541	ARLEVGLEQOALMLKMMSTLEADSILQALNTSPTLSQSPTGTDDSLGLGLOANQTSQ	2600	Db	3621	SSSIAQSIDISQDKURRHHVPOCCNKMPITADLVAPILRFLTEVGNHIMKDWLGSEV	3680
Qy	107	-----	106	Qy	129	SPTWSCI	135
Db	2601	LIITQSSVPMNLVCFNKFLSMQVHHVOLESLLQLWLTLSLNSSSGNKENGADIFLYNA	2660	Db	3681	NPLWTALLFLCHSGSTSGSHNLGAOOTSARSASLSAATTLTTOORTAIENATVAPFL	3740
Qy	107	-----	106	Qy	136	-----	135
Db	2661	NRIPVISLNOASITSFLTVLAWYNTLLRTWCLVLSLTLMTNQLNSGSSSAIGTOEST	2720	Db	3741	QCISCHPNKQKMAQVLCFQTSPOGNLPTSGNISGFIIRFLFLOLMEDEKVTMFLOS	3800
				Qy	136	-----	135

DB 3801 PCPLYKGRINATSHVIOHPMYGAGHKFRTLHLPVSTTSLDVLDRVSDTPSITAKLISQK 3860
QY 136 ----- 135
DB 3861 DDKEKNHEEKEKVAENGQDNYSVVASGLKSQSKRAVSATPPRPPRRGRTIPDKIG 3920
QY 136 ----- 135
DB 3921 STSGAANKIITVPFHLPHKLAGOPLPAEMTLAQLLTLLYDRKLPGVRSIDLTVKL 3980
QY 136 ----- 135
DB 3981 GSRVITDPSLSTDSYKRLHPKDHGDLASCPEDALTPGDCWDGILDESLLTCTPIQ 4040
QY 136 ----- 135
DB 4041 SPLQVAGMGGLIALERLPLVPEVIQVVSAPVVTSTTOEKPKDSQDFEWVTIEQSGL 4100
QY 136 ----- 135
DB 4101 VYEAPETVAEPPPIKSAVOTMSPAHSLAAGFLRLPGVAEVLKPKHQAQCLLRIV 4160
QY 136 --ITD----- 139
DB 4161 LGVTDGEGSHILOSPSANVPLTPLPHVLSLFTSTPLTTDDGVLLRRMALEIGALHLIL 4220
QY 140 -----TGFDLGV 147
DB 4221 VCLSALSHSPVNPSSVNOTEPOVSSHNPTSTEBQQLYWAQGTGFGTST 4272
RESULT 7
FAT2 HUMAN STANDARD; PRT; 4349 AA.
AC Q9NYQ8: O75091; QNSR7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protocadherin Fat 2 precursor (hFat2) (Multiple epidermal growth
factor-like domains 1).
GN FAT2 OR MEGF1 OR CDHF8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20202599; PubMed=10716726;
RA Wu Q., Maniatis T.;
RT "Large exons encoding multiple ectodomains are a characteristic
feature of protocadherin genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
RN [2]
RP SEQUENCE OF 3777-4349 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple
EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
RN [3]
RP SEQUENCE OF 4142-4349 FROM N.A.
RC TISSUE=Testis;
RA Poustka A., Wellenreuther R., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 32 CADHERIN DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.eib.ch/announce/>
or send an email to license@eib.ch).
EMBL; AF231022; AAF61928.1; -
EMBL; AB011535; BAA32463.1; -
EMBL; AL157443; CAB75663.1; -
Genew; HGNC:3596; FAT2.
MIM; 604269; -
HSP; P00740; IEDM.
InterPro; IPR002126; Cadherin.
InterPro; IPR000561; EGF-like.
Pfam; PF00028; cadherin; 33.
Pfam; PF00008; EGF; 2.
PRINTS; P00205; CADHERIN.
SMART; SM00112; CA; 32.
SMART; SM00181; EGF; 2.
SMART; SM00282; LamG; 1.
PROSITE; PS00232; CADHERIN_1; 14.
PROSITE; PS00268; CADHERIN_2; 32.
PROSITE; PS00022; EGF_1; 2.
PROSITE; PS01186; EGF_2; 1.
PROSITE; PS00025; LAM_G DOMAIN; 1.
Transmembrane; Glycoprotein; Repeat; EGF-like domain; Signal.
SIGNAL 1 18
CHAIN 19 4349 PROTOCADHERIN FAT 2.
DOMAIN 19 4048 EXTRACELLULAR (POTENTIAL).
TRANSMEM 4049 4069 POTENTIAL.
CYTOPLASMIC 4070 4349 CYTOPLASMIC (POTENTIAL).
DOMAIN 34 148 CADHERIN 1.
DOMAIN 149 256 CADHERIN 2.
DOMAIN 363 458 CADHERIN 3.
DOMAIN 459 564 CADHERIN 4.
DOMAIN 565 669 CADHERIN 5.
DOMAIN 716 820 CADHERIN 6.
DOMAIN 821 925 CADHERIN 7.
DOMAIN 926 1032 CADHERIN 8.
DOMAIN 1033 1137 CADHERIN 9.
DOMAIN 1138 1242 CADHERIN 10.
DOMAIN 1243 1346 CADHERIN 11.
DOMAIN 1350 1448 CADHERIN 12.
DOMAIN 1449 1555 CADHERIN 13.
DOMAIN 1556 1660 CADHERIN 14.
DOMAIN 1661 1758 CADHERIN 15.
DOMAIN 1759 1872 CADHERIN 16.
DOMAIN 1969 2070 CADHERIN 17.
DOMAIN 2071 2171 CADHERIN 18.
DOMAIN 2172 2272 CADHERIN 19.
DOMAIN 2273 2379 CADHERIN 20.
DOMAIN 2380 2481 CADHERIN 21.
DOMAIN 2482 2585 CADHERIN 22.
DOMAIN 2586 2691 CADHERIN 23.
DOMAIN 2692 2797 CADHERIN 24.
DOMAIN 2798 2906 CADHERIN 25.
DOMAIN 2907 3011 CADHERIN 26.
DOMAIN 3012 3113 CADHERIN 27.
DOMAIN 3114 3218 CADHERIN 28.
DOMAIN 3219 3321 CADHERIN 29.
DOMAIN 3322 3426 CADHERIN 30.
DOMAIN 3427 3531 CADHERIN 31.
DOMAIN 3532 3642 CADHERIN 32.
DOMAIN 3773 3944 LAMININ G-LIKE.
DOMAIN 3951 3984 EGF-LIKE 1.
DOMAIN 3989 4022 EGF-LIKE 2.
DOMAIN 3297 3302 POLY-SER.
CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 459 459 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).


```
FT CARBOHYD 627 627 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 655 655 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 789 789 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 996 996 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 1175 1175 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 1303 1303 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 1383 1383 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 1417 1417 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 1904 1904 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 1998 1998 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 2007 2007 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 2165 2165 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 2183 2183 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 2325 2325 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 2368 2368 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 2387 2387 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 2430 2430 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 2470 2470 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 2547 2547 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 2597 2597 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 2654 2654 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 3125 3125 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 3276 3276 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 3310 3310 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 3430 3430 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 3471 3471 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 3601 3601 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 3772 3772 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 3813 3813 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 3840 3840 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 3873 3873 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 3904 3904 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 3989 3989 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CONFLICT 4117 4117 P -> L (IN REF. 2) .
FT CONFLICT 4160 4160 E -> G (IN REF. 2) .
SQ SEQUENCE 4349 AA; 479383 MW; 6C050CA17AD398ED CRC64;

Query Match 27.8%; Score 215; DB 1; Length 4349;
Best Local Similarity 2.0%; Pred.No. 24;
Matches 88; Conservative 25; Mismatches 24; Indels 4198; Gaps 29;

Oy 5 ILH----- 7
    :||
Db 12 LHCATCEKPLEGILSSAHFTHSHYNATIYENSSPKTYVESFERKGIYLAEPQWAVRY 71
Oy 8 -----LEL--- 10
    :||
Db 72 RIISGDVANFKTEEYVVGNGFCFLRIRTKSSNTALLNREVRDSYTLIIQATEKTELEAL 131
Oy 11 ----- 10
Db 132 TRVVHILDNDLKLFLSPSPRYVTISEDMDPLKSPICKVTATDADLGQNAEFYAFNTRS 191
Oy 11 ----- 10
Db 192 EMFAIHPTSGVTVAGKLVNTHRGKHELQVLAVDNRMKI SEGNGFGSLAALVVHVEPALR 251
Oy 11 ----- 10
Db 252 KPPAIASVVVTPPDSNDGTTYATVLVDANSNGAEVESVEVGGDPGKFKAIKSYARSNE 311
Oy 11 --LVAV----- 14
    :||
Db 312 FSLVSVKDNIMWEYLHGFNLSQLQARSGBPYPFYQIRGPHLPPSKLSSLKFEKAVRVQL 371
Oy 15 ----- 14
Db 372 SEFSPGSRVVMVTPAFNLOVVLKPPSENVGFKLNARTGLITTTKLMDFDRAHYQL 431
Oy 15 ----- 14
Db 432 HIRTSPOQASTVVVIDIVDCNNHAPLNRSSYDGLDENIPPGTSVLAVTATDRDHGENG 491
```

```
Oy 15 -----GP----- 16
    :||
Db 492 YVTVSIAGPKALPESIDPYLGIISTSKPMDELMKRIYTFRVRASDNGSPRREKEVSIF 551
Oy 17 -----DV----- 18
    :||
Db 552 LOLRNLNDNOPMFEVNVCTGSIRODWPVGKSIIMTMSAIDVDELQNLKYEIVSGNELEYFD 611
Oy 19 -----FOAH----- 18
    :||
Db 612 LNHFGVISLKRPFINLTAGQPTSYSLKITASDCKNVASPTTLNITVVVKDPHFPEVPTCD 671
Oy 19 -----FOAH----- 22
    :||
Db 672 KTGVLTOFTKTLHFHIGLQNOESSDEFTSLSTYQINHYTPOFEDHFPQS:DVLESVPIN 731
Oy 23 ----- 22
    :||
Db 732 TPLARLAATDPDAGFNCKLVVVIADGNEEGCFDIELETGLTVAAPLDYEATNFILNVT 791
Oy 23 -----QEDTE----- 27
    :||
Db 792 VYDLGTQKSSMKLLTVNVKMDNDNAPRFPGGVQLTISEDTEVGTGTTIAELTTKDAISED 851
Oy 28 -----RVVL----- 31
    :||
Db 852 NGRVRYTLLSPTEKFSHLPLTGLVVTGHLDRSEPRYLKVEARDOPSKGHOLFVSVDL 911
Oy 28 -----RVVL----- 31
    :||
Db 912 IITLEDVNDNSPOCITEHNRUKLPEDLPPGTVLTFDASDPDLGPAGEVRVVLMDGANGT 971
Oy 32 ----- 31
    :||
Db 972 FRVDLMTGALLERBELDFERRAGYNLSLWASDGRPLARRTLCHVEVIVLDVNVNENLHPPH 1031
Oy 32 ----- 31
    :||
Db 1032 FASFVHQGOVQENSPSGTQVIWAAQDDSDGLDGELOYFLRAGTGLAAFSINQDTGMIOT 1091
Oy 32 ----- 31
    :||
Db 1092 LAPLDREFASYWLTVLAVDRGVSPLSSVTEVYIEVTANDNPPQMSOAVFYPSIOEDAP 1151
Oy 32 ----- 31
    :||
Db 1152 VCTSVLQLDANDPDSSSKGKLTFNITSGNYMGFFMIHPVTGLLSTAQOLDRENKDEHILE 1211
Oy 32 ----- 31
    :||
Db 1212 VTVLNNGEPLSKSTRVVVVGILDVNDNPPIFSHKLFNVRLPERLSPSPGVPYRLVASDL 1271
Oy 32 ----- 31
    :||
Db 1272 DEGLNGRVTSIEDSYBEAFSIDLVTVGVSSNSTFTAGEYNILTIKATDSGOPPLSASVR 1331
Oy 32 ----- 31
    :||
Db 1332 LHIEWIHPWRPSSIPLAFDETYYSFTVMTDPVNHVGVISVEGRPGLFWFNISGDKDM 1391
Oy 32 -----TNLNI----- 36
    :||
Db 1392 DFDIEKTTGSIVIARPLDTRRRSNYNLTVEVTDGSRTIATQVHIFMIANINHHRPQFLET 1451
Oy 37 -----GAELLR----- 42
    :||
Db 1452 RYEVVRPQDTPVEGVLLRQVAIDQDKGSLIYTHIGSQDPGSASLFOLDPPSGSVLTVCK 1511
Oy 43 ----- 42
    :||
Db 1512 LDLGSGPSOHTLTVMRDQEIPIKBNFVWVTHVEDGNLHPPRFTQLHYEASVPDTIAPG 1571
Oy 43 ----- 42
    :||
```

Db 1572 TELLQVRAMDADRGVNAEVHYSLKGNSEGFNINALLGIITLAQKLDQANHAPHTLTVK 1631
Qy 43 ----- 42
Db 1632 AEOGSPQWHDLATVIHVVYDSRSAPIFSKEVFEIPESIPVGSPIILLVSAMSPSEVT 1691
Qy 43 ----- 42
Db 1692 YELREGNKQGVFSMNSYSGLIISTQKLDHEKISSYQKIRGSNAGAFDTVMVVVDIIDE 1751
Qy 43 ----- 42
Db 1752 NDNAPMLKSTFFVQISEAAPLYSMINDKXNNPFVHASDSDKEANSLLYVKILEPEALK 1811
Qy 43 ----- 53
Db 1812 FFKIDPSMGTLTIVSEMDYESMPSFQFCVYVHDQGSPLFAPRAQVIIHVRDNDSPPR 1871
Qy 54 ----- 53
Db 1872 FSEQIYEVATVGPHPGMELMVRASEDESEVNSIKTGNADSAVTIHPVTGSIISVLNPA 1931
Qy 54 ----- 60
Db 1932 FLGLSRKLTIRASDGLYQDTALVKISLTQVLDKSLQFDQDVYAAVKENLQDRKALVILG 1991
Qy 61 ----- 60
Db 1992 AQGNHLNDTLYFLLNGTDMFHMVQSAQVLTQRCVAFDREQDTHELAVEVRDNRTPQV 2051
Qy 61 ----- 63
Db 2052 AOGLVRVSIEDVNDNPPKFKLPHYTYIIQDGTBFGDVLQVQSATDEDLGTNGAVTYEFAE 2111
Qy 64 ----- 63
Db 2112 DYTFRIDPYGLDISLKKPFQYQALNKYHLKVARDGGTSLQSEEEVLTVRNKSNPLF 2171
Qy 64 ----- 63
Db 2172 QSPYKVRVPENITLYPTILHTQARSPEGLRIYNIIVEEPMLFTTDFKTGVLTVTGPL 2231
Qy 64 ----- 67
Db 2232 DYESTKRVFTVRATDTALGSFSEATVEVLVEDVNDNPPPTFSQLVYTTISSEGLPAQTPV 2291
Qy 68 ----- 67
Db 2292 IQLLASQDSGRNRDVSQIVEDGSDVSKFQINGSTGEMSTVQELDYEAOQHVKVRA 2351
Qy 68 ----- 83
Db 2352 MDKGDPLTGELTVVNVVSDINDNPPEFRQPOYEANV--SELATCGHLVLKVQAIDPDSR 2409
Qy 84 ----- 83
Db 2410 DTSRLEYLILSGNDRHFFINSSSGIISMFNLCCKHLDSSYNLRVGASDGVFRATVPVI 2469
Qy 84 ----- 83
Db 2470 NTTNANKYSPEFQQLHYEAEIAENAMVGTVIDLAIQDKSGPYGTIDYTIINKLASEKF 2529
Qy 84 ----- 83
Db 2530 SINPGOIA TLQKLDRENSTERVIAIKVMARDGGGRVAFCTVKIILTENDNPPQPKASE 2589
Qy 84 ----- 91
Db 2590 YTVSIOQNVSKDSPVIOVLAYDADEGQNAVDTYSVNPEDLVKDVIEINPVTGVVVKRDSL 2649
Qy 92 ----- 91

Db 2650 VGLENOTLDFFIKAQDGGPPHMNSLVPVRLQVVPKKVSLPKFSEPLYTFSAPEDLPEGSE 2709
Qy 92 ----- 91
Db 2710 IGIVKAVAAQDPVIYSLVRGTTPESENKOGVFSLOPDTGVIKVRKPMHDESTKLYQIDVMA 2769
Qy 92 ----- 91
Db 2770 HCLQNTDVSLSVNIQVGDVNDNRVPVFEADPYKAVLTENMPVGTSVIOVTAIDKOTGRD 2829
Qy 92 ----- 91
Db 2830 GOVSRLSADPGSNVHELFAIDSEGWITTLQELDCETCTYHFHVAVYDHGOTIQLSSQ 2889
Qy 92 ----- 91
Db 2890 ALVQVSIITDENDNAPRFASEEYRGSVVENSEPCELVATUKTLADISEONRQVTCVITEG 2949
Qy 92 ----- 91
Db 2950 DPLQFGISQVGDWRISSRKTLDRHTAKYLLRVTASDGKFAQSVTVEIFVLVDVNDNSP 3009
Qy 92 ----- 97
Db 3010 QCSQLLYTGKVHEDVPPGHFILKVSATDLDTDNAQIITYSLHGPGAHEFKLDPHTGELTT 3069
Qy 98 ----- 104
Db 3070 LTALDRERKDVNLVAKATDGGRCQADITLHVEDVNDNAPRFPFSSHCAVAFDNTTVK 3129
Qy 105 ----- 104
Db 3130 TPVAVVPARDPQGANAOVVYSLPDSAEGHFSIDATTGVIRLEKPLQVRPOAPLELTVRA 3189
Qy 105 ----- 113
Db 3190 SDLGTBPLSTLGTVTVSVVGLDYLPFLNTEHSVQVPEDAPPGEVLQLATLTRPGAE 3249
Qy 114 ----- 119
Db 3250 KTGVRVYSGNEGRFLDARTGILYVNASLDPETSPKYFLSIECSRKSSSSLDVTTVMV 3309
Qy 120 ----- 119
Db 3310 NITDVNEHRPQPPQDPYPSTRVLENALVGDVILTVSATDEGPLNSDITYSLIGGNQLGHF 3369
Qy 120 ----- 119
Db 3370 TIHPKKGELQAKALDREQASSYSLKLRTDSCQPLHEDTDIAIQVADVNDNPPRFFQL 3429
Qy 120 ----- 119
Db 3430 NYSTTVQENSPIGSKVLQILSDPSDPENGPYPSPRITKGNNGSAPRVTPDGLVLTAEGL 3489
Qy 120 ----- 119
Db 3490 SRRAEWYQLOIQASDGPPLSSLTSVRVHVTEQSHYAPALPLEIFITVGEDEFQGM 3549
Qy 120 ----- 119
Db 3550 VGKIHATDRDPQDTLTYSLAEBETLGRHFSVGAPDGKIIAAOGLPRGHYSFNVTVSDGTF 3609
Qy 120 ----- 119
Db 3610 TTTAGVHVYVHVHQBALQOAMMGFYQLTPEELVSDHWRNLQRFSLKLDIKRANIHLA 3669
Qy 120 ----- 119
Db 3670 SLOPAEAVAGVDVLLVFECHSGTFFYEFQSLASTIITHSAKEMEHSVGVQMSAMPVPCOG 3729
Qy 120 ----- 119
Db 3730 PTCQGOICHNTVHLDPKVGPTYSTARLSILTPRHHLQRCSCNGTATRFSGQSVRYRAP 3789

```
OY 120 -----GVTQL----- 124
Db 3790 AARNWHIHYLKTLPQAILFTNETASVSLKASGVQPLEYHCLGFFYGNLSSQRHVND 3849
OY 125 ----- 124
Db 3850 HEWHSILVEEDASIRLMVDSMONTSLAVPENCRLRPERHLLGLILLHSSSNVSOGF 3909
OY 125 -----GGACSPWMS-- 133
Db 3910 EGCLDAVVNVEALDLPAGKTVAGLLETQALTOCCLHSDYCSQNTCLNGGKCS--WTHG 3967
OY 134 ----- 133
Db 3968 AGVVKCPQPFQSKHCEQRENCIFAPCLEGGTCILSPKASGNCNCPHYTGDRCEMEARG 4027
OY 134 -----CLIT----- 137
Db 4028 CSEGHCLVTEIQRGDWGQOELLITVAVAFIITVGLLFYCRCKSHKPVAMEPDL 4087
OY 138 ----- 137
Db 4088 ARSVGVDTOAMPALNPLSASNNLQNPESKASVPNELVTFGPNSKORPVVCSVPR 4147
OY 138 ----- 137
Db 4148 LPAAVPSHSDNEPVIKRTWSSEEMVPGAMVWPPTYSRNERWEYPHSEVTQGPLPSPA 4207
OY 138 ----- 137
Db 4208 HHRSTPVMPEPNGLYGGFPFPLEMENKRAPLPYRYSNQNLDMPSRPRRLVAPC 4267
OY 138 -----EDTGF----- 142
Db 4268 LNEYTAIYHSQFROGGCGPCLADGGYKGVGMRLSRAGPSYAVCEVAGLPGQOPRV 4327
OY 143 -----DLG 145
Db 4328 PPNYEGSDMVESDYG 4342

RESULT 8
ID RYR1_PIG STANDARD; PRT: 5035 AA.
AC P16360;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ryanodine receptor 1 (Skeletal muscle-type ryanodine receptor) (RYR1)
DE (RYR-1) (Skeletal muscle calcium release channel).
GN RYR1 OR CRC.
OS Sus acrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Norwegian Landrace; TISSUE=Skeletal muscle;
RX MEDLINE=93036581; PubMed=1329581;
RA Harbitz I., Kristensen T., Bosnes M., Kran S., Davies W.;
RT "DNA sequence of the skeletal muscle calcium release channel cDNA and
RT verification of the Arg615-->Cys615 mutation, associated with porcine
RT malignant hyperthermia, in Norwegian landrace pigs."
RL Anim. Genet. 23:395-402 (1992).
RN [2]
RP SEQUENCE OF 1129-2801 FROM N.A.
RN Brenig B.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1129-2643 FROM N.A.
RC STRAIN=German Landrace; TISSUE=Liver;
RX MEDLINE=94117003; PubMed=8288238;
```

```
RA Leeb T., Schmolzl S., Brem G., Brenig B.;
RT "Genomic organization of the porcine skeletal muscle ryanodine
RT receptor (RYR1) gene coding region 4624 to 7929.";
RL Genomics 18:349-354 (1993).
RN [4]
RP SEQUENCE OF 4785-5035 FROM N.A.
RX MEDLINE=91065640; PubMed=2174405;
RA Harbitz I., Chowdhary B., Thomsen P.D., Davies W., Kaufman U.,
RA Kran S., Gustavsen I., Christensen K., Hauge J.G.;
RT "Assignment of the porcine calcium release channel gene, a candidate
RT for the malignant hyperthermia locus, to the 6p11-->q21 segment of
RT chromosome 6.";
RL Genomics 8:243-248 (1990).
CC -!- FUNCTION: COMMUNICATION BETWEEN TRANSVERSE-TUBULES AND
CC SARCOPLASMIC RETICULUM. CONTRACTION OF SKELETAL MUSCLE IS
CC TRIGGERED BY RELEASE OF CALCIUM IONS FROM SR FOLLOWING
CC DEPOLARIZATION OF T-TUBULES.
CC -!- SUBUNIT: HOMOTETRAMER (POTENTIAL).
CC -!- MISCELLANEOUS: The calcium release channel is modulated by calcium
CC ions, magnesium ions, ATP and calmodulin.
CC -!- MISCELLANEOUS: The calcium release channel activity resides in the
CC C-terminal region while the remaining part of the protein
CC constitutes the 'foot' structure spanning the junctional gap
CC between the SR and the T-tubule. It is possible that the foot
CC structure interacts with the cytoplasmic region of the
CC dihydropyridine receptor.
CC -!- MISCELLANEOUS: Ryanodine is an alkaloid that binds to the Ca-
CC release channel in junctional SR and modulates its activity.
CC -!- SIMILARITY: BELONGS TO THE RYANODINE RECEPTOR FAMILY.
CC -!- SIMILARITY: CONTAINS 3 SPRY DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X62880; CAA44674.1; ALT_SEQ.
CC EMBL; X68247; CAA48318.1; -.
CC EMBL; X69465; CAA49225.1; -.
CC EMBL; M32501; AAA31022.1; -.
CC PIR; A37105; A37105.
CC InterPro; IPR000699; Ca-rel channel.
CC InterPro; IPR001682; Ca/Na_pore.
CC InterPro; IPR002048; EF-hand.
CC InterPro; IPR000636; M-channel_nlg.
CC InterPro; IPR003032; RYR.
CC InterPro; IPR001215; Ryanodn receptor.
CC InterPro; IPR003878; SPRY_domain.
CC InterPro; IPR003877; SPRY_receptor.
CC Pfam; PF00520; Ion_trans; 1.
CC Pfam; PF01365; RYDR_ITPR; 3.
CC Pfam; PF02026; RYR; 4.
CC Pfam; PF02815; MIR; 4.
CC PRINTS; PR00795; RYANODINER.
CC SMART; SM00472; MIR; 4.
CC SMART; SM00449; SPRY; 3.
KW Receptor; Transmembrane; Ionic channel; Calcium channel; Repeat;
KW Phosphorylation; Glycoprotein.
FT DOMAIN 1 3124
FT TRANSMEM 3125 3145 M' (POTENTIAL).
FT TRANSMEM 3189 3207 M'' (POTENTIAL).
FT TRANSMEM 3980 3999 M1 (POTENTIAL).
FT TRANSMEM 4018 4036 M2 (POTENTIAL).
FT TRANSMEM 4274 4297 M3 (POTENTIAL).
FT TRANSMEM 4339 4359 M4 (POTENTIAL).
FT TRANSMEM 4557 4578 M5 (POTENTIAL).
FT TRANSMEM 4646 4669 M6 (POTENTIAL).
FT TRANSMEM 4787 4807 M7 (POTENTIAL).
```

FT	TRANSMEM	4835	4854	M8 (POTENTIAL).
FT	TRANSMEM	4877	4896	M9 (POTENTIAL).
FT	TRANSMEM	4921	4935	M10 (POTENTIAL).
FT	DOMAIN	659	797	SPRY 1.
FT	DOMAIN	1085	1208	SPRY 2.
FT	DOMAIN	1430	1570	SPRY 3.
FT	DOMAIN	1874	1925	GLU-RICH (ACIDIC).
FT	DOMAIN	4458	4526	PRO-RICH.
FT	DOMAIN	842	2960	6 X APPROXIMATE REPEATS.
FT	REPEAT	842	955	1.
FT	REPEAT	956	1069	2.
FT	REPEAT	1345	1360	3. (INCOMPLETE).
FT	REPEAT	1373	1388	4. (INCOMPLETE).
FT	REPEAT	2727	2846	5.
FT	REPEAT	2847	2960	6.
FT	MOD_RES	2844	2844	PHOSPHORYLATION (BY PKA AND PKG) (BY SIMILARITY).
FT	MOD_RES	3947	3947	PHOSPHORYLATION (POTENTIAL).
FT	MOD_RES	4320	4320	PHOSPHORYLATION (POTENTIAL).
FT	CARBOHYD	3467	3467	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3475	3475	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3904	3904	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3945	3945	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4144	4144	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4862	4862	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	2092	2092	A -> P (IN REF. 2 AND 3).
FT	SEQUENCE	5035	5035	AA; 565317 MW; E00613F2027B94A CRC64;

Query Match 27.4%; Score 212; DB 1; Length 5035;

Best Local Similarity 2.6%; Pred. No. 51;

Matches 83; Conservative 32; Mismatches 33; Indels 3096; Gaps 27;

QY	1	AAGGILHL	8
DB	556	ASSGILEVLCVLESPEVLNIOENHIKSIISLLDKHGRNKHVLDVLCSCVCGNAVR	615
QY	9	-----ELLV-----AVGPDVFOAQHOEDERY-----	29
DB	616	SNQDLITENLLPGRELLQTLNLYVTSIRENIFVGRAEGTQYKSYFVEMVDEVVPL	675
QY	30	-----VLT-----	32
DB	676	TAQATHLRVGWALTGEGSYYPGGGEGGNGVDDLSYGFDDLHWTGHVPRLVTSQG	735
QY	33	-----NLN-----	35
DB	736	HLLAPEDVVSCLDLSVPSISFRINGCPVQGVFAFNGLFFPVVVSAGVKVRFLLGG	795
QY	36	-----	35
DB	796	RHGEFKFLUPPGYAPCHEAVLPRERLRLEPIKEYRREGPRGPHLVGPSRCLSHTDVPCP	855
QY	36	-----	35
DB	856	VDTVQIVLPHLERIREKLAENIHELWALTREOGWTYGPVRDDNKRHLRHLPCLVDFHSLPE	915
QY	36	-----	35
DB	916	PERNYLQMSGETLKTLLALGCHVGMADEADNLRKTKLPKTYMMSNGYKPAPLDLSHV	975
QY	36	-----	35
DB	976	RLTPAQTTLVDRLAENGHNWWARDVAQGSYSYAVQDIPARRNPRLVPYRLDLEATKRSN	1035
QY	36	-----	35
DB	1036	RDSLQCAVRTLGGYNIETPPDQEPQSVESQSRWDRVIRFRAEKSYAVQSGRWYFEFAV	1095
QY	36	-----IGAEILLRDP-----	44
DB	1096	TTGEMRVGWARPELRPDVELGADELAYVFNHGRGWRHLGSELFGRPWQSGDVVCCMIDL	1155
QY	45	-----	44

```
Db 2236 FLCYFCRISRONQSMFDHLSYLLSENSIGLGMQSGTFLDVAAASVIDNNELALALOEQD 2295
Qy 94 -----
Db 2296 LEKVSVYLAGCGLQSCPMLLAKYDPDIGNPCGGERYLDLFLRFAVFVNGESVEENANVVV 2355
Qy 94 -----
Db 2356 RLLIRKPECFGPALRGEGSGLLATIBEARISEDPA RDGVRDRRREHFGEEPPBEEN 2415
Qy 96 ---CHA---
|||
|||
Db 2416 RVHLGHAIMSYAALIDLLRCAPEMHUIQAGKGEALRILRLSLVPLDVLGIIISLPL 2475
Qy 102 -----
Db 2476 QIPTLGKDALVOPKMSASFPVDHKASVFLDRVYGIENQDFLLHVLVDGFLPDMRAA 2535
Qy 102 -----
Db 2536 SLDTATFTTETMALNRYLCLAVLPLITKCAPLPAGTEHRAIMVDSMLHTVYRLSGRS 2595
Qy 102 -----
Db 2596 LTKAQRDIECLMALCRYIRPSMLQHLRLVFDVPILNEFAKMPKLKLTNNHYERCWKY 2655
Qy 102 -----
Db 2656 YCLPTGWNFGVTSEELHLTRKLPWGIFDSLAHKKYDPELYRMAMPCLCAIAGALPDY 2715
Qy 114 -----
Db 2716 VDASYSSAKKATVDAEGNPDPRPVETLNVIIPEKLDSEFINKEFAEYTHEKWAEDKIQNN 2775
Qy 116 -----
Db 2776 WSYGENIDEELKTHPLRPYKTFSEKDEIYRWPIKESLKAMIAEWETIEKAREGEBEKT 2835
Qy 116 -----
Db 2836 EKKTRKISQSAQTYDAREGYNQPPDLSGVTLRSLOMAEQLAENYHNTWGRKKQEL 2895
Qy 116 -----
Db 2896 EAKGGTHPLLPVDTLTAKKARDREKAQELLKFLQMGYAVTRGLKDMELDTSSIEKR 2955
Qy 116 -----
Db 2956 FAFGLQQLLRWMDISQEFIAHLEAVSSGRVKEKSPHEQETKFFAKILLPLINQVFTNHC 3015
Qy 116 -----
Db 3016 LYFLSTPAKVLGSGGHASNKEMITSLFCKLAALVRHVSILFGDPAVAVNCLHILARS 3075
Qy 116 -----
Db 3076 LDARTVMKSGPEIVKAGLRSFESASEIEKMNVLRLGKVSQARTQVKGVGQNLTYTV 3135
Qy 124 -----
Db 3136 ALLPVLTLTFOHIAHQFGDDVILDVQVSVTRTLCISYSLGTTNPNVVEKLRPALGEC 3195
Qy 124 -----
Db 3196 ARLAAMPVAFLEPQLENEYNACSVYTKSPREBAILGLPNSVEEMCPDIPVLERLMAIDG 3255
Qy 126 G-----
Db 3256 GLAESGARYTEMPHVIEITLPMCLSYLPRWMERGPEAPPALPAGAPPPTAVTSDHLS 3315
Qy 127 -----
Db 3316 LLGNILRIIVNNLGIDEASMMKRLAVFAQPIVSRARPELLHSHFPTIGRLRKRAKVVA 3375
```

RESULT 9

```
FAT2_RAT
ID FAT2_RAT STANDARD; PRT; 4351 AA.
AC O88277;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protocadherin Fat 2 precursor (Multiple epidermal growth factor-like domains 1).
GN FAT2 OR MEGF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
CC -1- SIMILARITY: CONTAINS 33 CADHERIN DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011527; BAA32458.1; -
DR HSPR; P08709; IBF9.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
```

Query Match	Best Local Similarity	Score	DB 1	Length	4351
Matches	78	Conservative	35	Mismatches	35
Indels	4077	Gaps			
FT CARBOHYD 2325	2325	2325	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 2368	2368	2368	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 2387	2387	2387	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 2430	2430	2430	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 2470	2470	2470	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 2547	2547	2547	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 2597	2597	2597	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 3127	3127	3127	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 3278	3278	3278	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 3312	3312	3312	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 3432	3432	3432	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 3603	3603	3603	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 3770	3770	3770	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 3774	3774	3774	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 3815	3815	3815	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 3842	3842	3842	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 3875	3875	3875	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 3906	3906	3906	N-LINKED (GLCNAC. .)	(POTENTIAL)	
FT CARBOHYD 3991	3991	3991	N-LINKED (GLCNAC. .)	(POTENTIAL)	
SQ SEQUENCE 4351 AA	480644 MW	C564E7F98B3888 CRC64			
Query Match	27.3%	Score 211	DB 1	Length 4351	
Best Local Similarity	1.8%	Pred. No. 37			
Matches	78	Conservative	35	Mismatches	35
Indels	4077	Gaps			
QY 1 AAG	-----	-----	-----	-----	3
DB 78 AAGVFKTEHVGVNFCFLRIRTKSNTALLNREVRDSYTLIVQASDKSLEALQVVVH	137				
QY 4	-----	-----	-----	-----	3
DB 138 ILQNDLKLPLSPSPSYRVITSEDRPLKSPICKVTATDADLQNAEFYAFNARSEVP	197				
QY 4	-----	-----	-----	-----	3
DB 198 PTSGVTVVAGKLVNTRRGKYLQVLAVDRMRKISEGNFGNLSLVIRVEPVHRKPP	257				
QY 4	-----	-----	-----	-----	3
DB 258 LVVLNPPGDEGDIYAIIVTVDTNGSGAEVDSLEVVGDPGKFKVLRSYAOGNEFN	317				
QY 4	-----	-----	-----	-----	3
DB 318 RDINWAEHPHGFNISLQTHSNRFPFPHSIIRAFHLPMSKLANLRPEKAVRVKLS	377				
QY 4	-----	-----	-----	-----	3
DB 378 GSRVALVKVTTALPNLRYSLKPSRNTAFKLNARTGLITTTKLVDHFQEQVOLHV	437				
QY 4	-----	-----	-----	-----	3
DB 438 QOATTVIIVDCNNHAPVFNRRSYEGTLDENIPGTSVLTVTATDQDHGNGHIT	497				
QY 4	-----	-----	-----	-----	3
DB 498 AGPKAVPFSIDPLGLVISTTKPMYELMKRIYTPRVRASDWGSPFROEKEVS	557				
QY 4	-----	-----	-----	-----	3
DB 558 NDQPMFEVNCVTSRLQRDPVGVKSIMAVSAIDMDELQNLKYEIVSGNEBQDY	617				
QY 4	-----	-----	-----	-----	6
DB 618 VISLKRSPMNLTAVRPTIYSLKITASDGKNYASPTTLKVTVVKDPHSEVP	677				
QY 7	-----	-----	-----	-----	6
DB 678 HITKTLQASAGLOSQELGEEFTSLSNYQINHHSPQEDHPQSIDILSOVPINT	737				
QY 7	-----	-----	-----	-----	6
DB 738 AATDPDTGPHGKLVYVISDGBEGCFDIELETGLLWAAALDYETTSFYVLNV	797				

Qy 7 ----- 6
Db 798 PPKSSMKLLTVTKWQNDNPPRPPGGYQLTISEDTEVGTITIAELKTEDADSEDNRVRV 857
Qy 7 ----- 6
Db 858 TLLTPTEKSLHPFTGELVVTGHLDRSESOYLKAEARDOPTKGHOLFVSTDLIVTLIED 917
Qy 7 ----- 6
Db 918 INDNPPQICITEHRRRLKVPEDMPLGTVLTFDASDDPLDGPAGEVKYILVEDAHGTFQVHPM 977
Qy 7 ----- 12
Db 978 TCALSLEKELDFERRAGYNLSFWASDSKPLSRRTLCHVEVLVMDVNNLHSPHSSFVY 1037
Qy 13 ----- 24
Db 1038 OQVOENSPAGTPVMVTAQDDSDGLDELQYFLRAGTGLETFSINQDTGMLTAPLADR 1097
Qy 25 ----- 24
Db 1098 EFTPYNLTVLAVDRGSVPLSAVTEVYIEVTDINDNISMRPVPYPSVLEDAPLGTSVL 1157
Qy 25 ----- 24
Db 1158 QLEAWDPDSSOGKLTFNLTSGNLHGFIHVHPTGTLTTAKQLDRENKDEYVLEVTVQDN 1217
Qy 25 ----- 24
Db 1218 GDSLRSTSRVVVCIIDVNDNPPMFHKLFNVRLSERLSPSPVPYRLVASDPDEGLNG 1277
Qy 25 ----- 24
Db 1278 SVTYSIESDEESFRIDPVTGVVSSSTFAAGEYNILTIKATDSCQPALSTSVRLHIEWI 1337
Qy 25 ----- 24
Db 1338 POPRSSIPLSDSESYSTYMETDPVNMHVGVISVEGRPGLFWFHI SDGDKMDFDIEK 1397
Qy 25 ----- 35
Db 1398 TTGSIVARPLDTRRKSSVNLTVETDGFHTIATQVHFMIANINHRPQFLQDHYEIRV 1457
Qy 36 ----- 42
Db 1458 PQDTLPGVLELRVQATDQDHGKGLIYTLSSODPGSANLFOLDPSSGVLVTVTGTLHLHG 1517
Qy 43 ----- 42
Db 1518 PSQHILTVMVRQEMPIKRNFWVTIHVEDGNLHSPHFTQLRYEANVPDTPAGTTELQV 1577
Qy 43 ----- 42
Db 1578 RAVDADRGAEBIHYFLKNSDGFNIDSLIGIITVAORLYHVHLTRHALTVKAEDOGS 1637
Qy 43 ----- 42
Db 1638 PRRHDLVVIHVPSSAPVFSKDEYFIEIPESVPIGSPILLLSAGSSSEVTVELREG 1697
Qy 43 ----- 42
Db 1698 NKDSVFSMNSYSGLISTQKRLDHEKVPSYRLRIRGSNMAGVTEVALVYIIDENDNPPA 1757
Qy 43 ----- 44
Db 1758 FGKPTFLGHISEAAPHLSLILGEDNSPLVVRASDSDREANSLLVYKILEPEALKFKPIDP 1817
Qy 45 SLGA ----- 52
Db 1818 SMGTLTTTSEDFETPLPQFNIYVHDOGTPLFAPRSKAKYIIHVRDNDSPRPFSEQIY 1877
Qy 53 ----- 52

Db 1878 EVAVVEPIHFQMGLLTVQABEDNDSRVTSYIKTSNADEAVTIHPTTGOISVVNPATLRFQ 1937
Qy 53 ----- 60
Db 1938 KFSIRASDGLYHDTAVVVKISLTVLQKSLQFDQDQVYRARTENTPHRKALVILGVGHNL 1997
Qy 61 ----- 60
Db 1998 NDTLSYFLNGTDLFHMIESAGVLQTRGGTFDRQQDTHVAVEVRDNRVORVAQALVR 2057
Qy 61 ----- 63
Db 2058 VSEDVNDNIPEFOHLPYTVIOQGTPEGDVLFQVSAATDKDLGANGSVTYGAEADYAYFR 2117
Qy 64 ----- 63
Db 2118 IDPVYGDISLKKPPDYQALNKYHLVRIARSGIPPLQTEVEHVTVRNKSNPLFQSPYK 2177
Qy 64 ----- 63
Db 2178 VKPENITLYTPILHTOARSPGRLIYNIVEBEPMLFTTDFPKTVLTVTGPDYESKN 2237
Qy 64 ----- 67
Db 2238 KHVETRATDTALGSPSEATVEVLVEDINDNPPTFSQLVYTVTSVSEGPSAQTPIQLLAS 2297
Qy 68 ----- 67
Db 2298 DQSGQNDVSYQIVEDGSDVSKFRINGSTGEIFTQELDYETHQHPRVKVRAMDKGP 2357
Qy 68 ----- 67
Db 2358 PLTGETLVVNVSDINDNPPKPREPOYEANVSELATCGHLVLVKQALDPDIDGTSRLEYL 2417
Qy 68 ----- 80
Db 2418 ILSGNQDRHFSINSTSGIISMFNLCKQLDSSYNLRVGSQVPRATVPVYINTTNANKY 2477
Qy 81 ----- 80
Db 2478 SPEFQNVYBAELAENAKVKVIELLAIDKDSQPGYTVDYTIINKLAGERFFINPRQOI 2537
Qy 81 ----- 80
Db 2538 TTLQKLDRENSTERVIAIKVMARDGGKVAFCVTKIILTDENDNAPQFKASGYTVSIPSN 2597
Qy 81 ----- 80
Db 2598 VSRDSPIIQVLAYDADEGRNADVTYSVDSTEDLAEIIEVNPITGVVKKVGLVLENRA 2657
Qy 81 ----- 80
Db 2658 VDFNIKADGGGPPHWDLSLVPRLQVVPNEIPLPKFSEPLTYFSAPEDLPEGSEIGSVKAV 2717
Qy 81 ----- 80
Db 2718 AAQDPILYSLVQGTTPESNSDDVPSLDQDTGVLKVRKAMDHSTKMYQIDLMAHCPHEDT 2777
Qy 81 ----- 80
Db 2778 DLVLSVSVSIQVEDVNDNRVPFEADPYKAFLTENMPGTTVIOVTANDQDTGSDGOVSyr 2837
Qy 81 ----- 87
Db 2838 LSVPEGNIHELFAVDSESGWITTLQELDCETOOTYRFFVVAFDHGQTILQSSQALVEVS 2897
Qy 88 ----- 89
Db 2898 ITDENDNPPRFASEDYRGSVVENNEPGEVLVATLKLADVSDQNRQVTCYITEGDPLQCF 2957
Qy 90 ----- 89

```
Db 2958 SISQGVDEWIRSRKTLDRHIAKYLLRVKTASDGKFAQSVPEVFVVDINDNSPQCQOLL 3017
Qy 90 -----EDTDPRCH----- 97
Db 3018 YTGKVRDVTGPHFKLVKSAIDVDMOTNAQITYSLHGPAGAEFKLDPHTGELTTLTVLDR 3077
Qy 98 ----- 97
Db 3078 ERKDVYNLVAKATDGGGSCQAEVTLHIEDVNDNAPRFFPSHCDVAVFDNTTKTPVAVV 3137
Qy 98 -----ADL----- 100
Db 3138 FARDPOGANAQVYVSLTDSADGQFSIDATSGVIRLEKPLQVRASSAVELTVRASDLGTP 3197
Qy 101 ----- 100
Db 3198 IPLSTLGTVTVVVGLGEDYLPFIFLNAEHSHTQVPEDAPIDMEVLHLATLTPRGSEKTYGHI 3257
Qy 101 -----VLYI----- 104
Db 3258 TGGNEQCKFRDLAHTGILYVNGSLDFETNPKYFLSIECSRKSSSLSDVTTIVINVTDNV 3317
Qy 105 ----- 104
Db 3318 EHHPRFTHDLYTVRVLENVAVGDVILTVSASDDGPNVSAITYSLVGNQGLGHTINPKK 3377
Qy 105 ----- 104
Db 3378 GKLOVAKALDWEOTPSYSLRLRATDSGOPPLHEDTEVAVEVDVNDNPPRFFQLNYSTSV 3437
Qy 105 ----- 104
Db 3438 QENSPIGIKVLQLLDDPSPONGPPYFFRITEGNTGSFVRVTPDGLVLTAAASLKKARE 3497
Qy 105 ----- 104
Db 3498 WYQLHIEVSDGLPPLSSSTLVRVQVTEQSRYPSTPLLEISITKGEFQGMIGIHA 3557
Qy 105 ----- 104
Db 3558 TDRPDQTLTYSLEQEGGLDRYFTVGASDGKIIASQGLPHGRYSFNVTVSDGFTTTTGV 3617
Qy 105 ----- 104
Db 3618 HVHVHMEPEVPOAVNVLGFHOLTPEELVSDHHRNLFQFLNLLDVYKRAIHLASLOPAE 3677
Qy 105 -----TRFQLE----- 110
Db 3678 VTAGVDVLLVERHSGTSYDLQELASAIHVSREIHSVGIRMSALPVVPCQSQSQDQ 3737
Qy 111 ----- 110
Db 3738 TCQETVLEPRVGSYSTARLSILTTPRHILGRNCSCNGTTLRFSGOSYVQYRPLEAQNWQ 3797
Qy 111 ----- 110
Db 3798 IHFYLKTLQWALLMFTNETASISLKLANGFSLHYHCPGGFYGNLSSRYPVNDGQWHS 3857
Qy 111 -----LPDGNRQVRGVTL----- 124
Db 3858 LLEERDTSVHLVDITDNASLVIPEECQGLRTERQLLLGLVPSNPSSNVSLGFEGCLDA 3917
Qy 125 -----GGACSP----- 130
Db 3918 VVNGERLELLGRKKWEGRLWALSOCCWPGTACSQSPCLNGGSCSPALGSGYLCRCP 3977
Qy 131 ----- 130
Db 3978 PPSGRNCELGRNCTSAQCQEGGTCVSSPEGTSCNCPHYTGDRCEMARGCGGHCLI 4037
Qy 131 ----- 130
Db 4038 TPEIKRGDWGQOEFLVITVALPLVVIATVGLLLYCRRRKSHKPVMTMEDPDLARSIGVDT 4097
```

```
Qy 131 ----- 130
Db 4098 QASPAIELOPLNTSSNNLNQPEPSKTSVPNELVTGPPSSKORPMVCSVPPRLPPAAVSS 4157
Qy 131 -----TWS-----CL 135
Db 4158 HPGHEPIIKRTWSGEELVYPSGAAVWPPTYSRKKHWEYPHPETMOGTLPPSPRRHVGPAV 4217
Qy 136 ITEDTGF-----DL----- 144
Db 4218 MPDPTGLYGGFFPPELENKRAPLPPTYGNQNLDELMPRPSPREHLLAPCLNEYTAIS 4277
Qy 145 -----GVTI 148
Db 4278 YHSQLPROGGGGPCLAEAGYKGVSM 4302

RESULT 10
RYR1 HUMAN
ID RYR1 HUMAN STANDARD; PRT: 5038 AA.
AC P21817; Q16314; Q9NPX1; Q9PLU4;
DT 01-MAY-1991 (Rel. 18, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ryanodine receptor 1 (Skeletal muscle-type ryanodine receptor) (RYR1)
DE (RYR-1) (Skeletal muscle calcium release channel).
GN RYR1 OR RYDR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A. (ISOFORM 2), AND PARTIAL SEQUENCE.
RX TISSUE=Skeletal muscle;
RX MEDLINE=90130482; PubMed=2298749;
RA Zorzato F., Fujii J., Otsu K., Phillips M.S., Green N.M., Lai F.A.,
RA Meisner G., MacLennan D.H.;
RT "Molecular cloning of cDNA encoding human and rabbit forms of the
RT Ca2+ release channel (ryanodine receptor) of skeletal muscle
RT sarcoplasmic reticulum.";
RL J. Biol. Chem. 265:2244-2256(1990).
RN [2]
RP REVISIONS TO 2324; 2840 AND 3380, VARIANT MH ARG-248, AND VARIANTS
RP CYS-471; LEU-1787 AND CYS-2060.
RC TISSUE=Muscle;
RX MEDLINE=92372020; PubMed=1354642;
RA Gillard E.F., Otsu K., Fujii J., Duff C.L., de Leon S., Khanna V.K.,
RA Britt B.A., Morton R.G., McLennan D.H.;
RT "Polymorphisms and deduced amino acid substitutions in the coding
RT sequence of the ryanodine receptor (RYR1) gene in individuals with
RT malignant hyperthermia.";
RL Genomics 13:1247-1254(1992).
RN [3]
RP REVISIONS TO 1365-1368, VARIANT CCD H-2435, AND ALTERNATIVE SPLICING.
RC TISSUE=Muscle;
RX MEDLINE=94035117; PubMed=8220422;
RA Zhang Y., Chen H.S., Khanna V.K., de Leon S., Phillips M.S.,
RA Schappert K., Britt B.A., Brownell A.K.W., McLennan D.H.;
RT "A mutation in the human ryanodine receptor gene associated with
RT central core disease.";
RL Nat. Genet. 5:46-50(1993).
RN [4]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=95299657; PubMed=8661021;
RA Phillips M.S., Fujii J., Khanna V.K., de Leon S., Yokobata K.,
RA de Jong P.J., McLennan D.H.;
RT "The structural organization of the human skeletal muscle ryanodine
RT receptor (RYR1) gene.";
RL Genomics 34:24-41(1996).
RN [5]
RP SEQUENCE OF 598-722 FROM N.A.
RP TISSUE=Skeletal muscle;
```


RX MEDLINE=92347887; PubMed=1639409;
RA Otsu K., Phillips M.S., Khanna V.K., de Leon S., McLennan D.H.;
RT "Refinement of diagnostic assays for a probable causal mutation for
RL porcine and human malignant hyperthermia.";
RN Genomics 13:835-837(1992).
[16]
RP SEQUENCE OF 3339-5038 FROM N.A. (ISOFORM 2).
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RA Danganan L., Exler A., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Krommiller B., Arellano A., Saunders C., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
[17]
RP SEQUENCE OF 4696-4974 FROM N.A.
RC TISSUE=Myometrium;
RX MEDLINE=96032536; PubMed=7556644;
RA Lynn S., Morgan J.M., Lamb H.K., Meissner G., Gillespie J.I.;
RT "Isolation and partial cloning of ryanodine-sensitive Ca²⁺ release
RL channel protein isoforms from human myometrial smooth muscle.";
RN FEBS Lett. 372:6-12(1995).
[18]
RP TISSUE SPECIFICITY.
RX MEDLINE=98268728; PubMed=9607712;
RA Martin C., Chapman K.E., Seckl J.R., Ashley R.H.;
RT "Partial cloning and differential expression of ryanodine
RL receptor/calcium-release channel genes in human tissues including
the hippocampus and cerebellum.";
RN Neuroscience 95:205-216(1998).
[19]
RP S-NITROSYLATION OF CYS-3635.
RX MEDLINE=21457268; PubMed=11562475;
RA Sun J., Xin C., Eu J.P., Stampler J.S., Meissner G.;
RT "Cysteine-3635 is responsible for skeletal muscle ryanodine receptor
RL modulation by NO.";
RN Proc. Natl. Acad. Sci. U.S.A. 98:11158-11162(2001).
[10]
RP VARIANT MH CYS-614.
RX MEDLINE=92128959; PubMed=1774074;
RA Gillard E.F., Otsu K., Fujii J., Khanna V.K., de Leon S.,
RA Derdemezi J., Britt B.A., Duff C.B., Worton R.G., McLennan D.H.;
RT "A substitution of cysteine for arginine 614 in the ryanodine
RL receptor is potentially causative of human malignant hyperthermia.";
RN Genomics 11:751-755(1991).
[11]
RP VARIANTS CCD CYS-163 AND MET-403.
RX MEDLINE=94035118; PubMed=8220423;
RA Quane K.A., Healy J.M.S., Keating K.E., Manning B.M., Couch F.J.,
RA Palmucci L.M., Doriguzzi C., Fagerlund T.H., Berg K., Ordling H.,
RA Bendixen D., Mortier W., Linz U., Muller C.R., McCarthy T.V.;
RT "Mutations in the ryanodine receptor gene in central core disease and
RL malignant hyperthermia.";
RN Nat. Genet. 5:51-55(1993).
[12]
RP VARIANT MH SER-522.
RX MEDLINE=95130087; PubMed=7829078;
RA Quane K.A., Keating K.E., Healy J.M.S., Manning B.M.,
RA Krivosic-Horber R., Krivosic I., Monnier N., Lunardi J.,
RA McCarthy T.V.;
RT "Mutation screening of the RYR1 gene in malignant hyperthermia:
RL detection of a novel Tyr to Ser mutation in a pedigree with
associated central cores.";
RN Genomics 23:236-239(1994).
[13]
RP VARIANT MH ARG-341.
RX MEDLINE=94282042; PubMed=8012359;
RA Quane K.A., Keating K.E., Manning B.M., Healy J.M.S., Monstieus K.,
RA Heffron J.J.A., Lehane M., Heytens L., Krivosic-Horber R., Adnet P.,
RA Ellis F.R., Monnier N., Lunardi J., McCarthy T.V.;
RT "Detection of a novel common mutation in the ryanodine receptor gene

in malignant hyperthermia: implications for diagnosis and
heterogeneity studies.";
RN Hum. Mol. Genet. 3:471-476(1994).
[14]
RP VARIANT MH ARG-2434.
RX MEDLINE=95152512; PubMed=7849712;
RA Keating K.E., Quane K.A., Manning B.M., Lehane M., Hartung E.,
RA Censier K., Urwyler A., Klausnitzer M., Muller C.R., Heffron J.J.A.,
RA McCarthy T.V.;
RT "Detection of a novel RYR1 mutation in four malignant hyperthermia
RL pedigrees.";
RN Hum. Mol. Genet. 3:1855-1858(1994).
[15]
RP VARIANT MH ARG-2434.
RX MEDLINE=95187158; PubMed=7881417;
RA Phillips M.S., Khanna V.K., de Leon S., Frodis W., Britt B.A.,
RA McLennan D.H.;
RT "The substitution of Arg for Gly2433 in the human skeletal muscle
RL ryanodine receptor is associated with malignant hyperthermia.";
RN Hum. Mol. Genet. 3:2181-2186(1994).
[16]
RP VARIANT MH CYS-614.
RC TISSUE=Blood;
RX MEDLINE=95271229; PubMed=7751854;
RA Moroni I., Gonano E.F., Comi G.P., Tegazzin V., Prella A., Bordini A.,
RA Bresolin N., Scarlato G.;
RT "Ryanodine receptor gene point mutation and malignant hyperthermia
RL susceptibility.";
RN J. Neurol. 242:127-133(1995).
[17]
RP VARIANT MH ARG-35.
RX MEDLINE=97219028; PubMed=9066328;
RA Lynch P.J., Krivosic-Horber R., Reyford H., Monnier N., Quane K.A.,
RA Adnet P., Haudecoeur G., Krivosic I., McCarthy T.V., Lunardi J.;
RT "Identification of heterozygous and homozygous individuals with the
RL novel RYR1 mutation Cys35Arg in a large kindred.";
RN Anesthesiology 86:620-626(1997).
[18]
RP VARIANT MH LEU-614.
RX MEDLINE=98051290; PubMed=9389851;
RA Quane K.A., Ordling H., Keating K.E., Manning B.M., Heine R.,
RA Bendixen D., Berg K., Krivosic-Horber R., Lehmann-Horn F.,
RA Fagerlund T.H., McCarthy T.V.;
RT "Detection of a novel mutation at amino acid position 614 in the
RL ryanodine receptor in malignant hyperthermia.";
RN Br. J. Anaesth. 79:332-337(1997).
[19]
RP VARIANT MH TRP-552.
RX MEDLINE=97284075; PubMed=9138151;
RA Keating K.E., Giblin L., Lynch P.J., Quane K.A., Lehane M.,
RA Heffron J.J.A., McCarthy T.V.;
RT "Detection of a novel mutation in the ryanodine receptor gene in an
RL Irish malignant hyperthermia pedigree: correlation of the IVCT
response with the affected and unaffected haplotypes.";
RN J. Med. Genet. 34:291-296(1997).
[20]
RP VARIANTS MH CYS-2163; HIS-2163; MET-2168 AND MET-2206.
RX MEDLINE=98163444; PubMed=9497245;
RA Manning B.M., Quane K.A., Ordling H., Urwyler A., Tegazzin V.,
RA Lehane M., O'Halloran J., Hartung E., Giblin L.M., Lynch P.J.,
RA Vaughan P., Censier K., Bendixen D., Comi G.P., Heytens L.,
RA Monstieus K., Fagerlund T.H., Wolz W., Heffron J.J.A., Muller C.R.,
RA McCarthy T.V.;
RT "Identification of novel mutations in the ryanodine-receptor gene
RL (RYR1) in malignant hyperthermia: genotype-phenotype correlation.";
RN Am. J. Hum. Genet. 62:595-609(1998).
[21]
RP VARIANTS MH CYS-2459 AND HIS-2459.
RX MEDLINE=98111378; PubMed=9450902;
RA Manning B.M., Quane K.A., Lynch P.J., Urwyler A., Tegazzin V.,
RA Krivosic-Horber R., Censier K., Comi G.P., Adnet P., Wolz W.,
RA Lunardi J., Muller C.R., McCarthy T.V.;
RT "Novel mutations at a CpG dinucleotide in the ryanodine receptor in

```
RT malignant hyperthermia. ";
Query Match 27.2%; Score 210; DB 1; Length 5038;
Best Local Similarity 2.5%; Pred. No. 64;
Matches 81; Conservative 33; Mismatches 34; Indels 3101; Gaps 26;

Qy 1 AAGGILHL----- 8
Db 555 ASSGILEVLCVLTIESPEVLNIIQENHIKSIISLLDKHGRNHKVLVDVLCVLCVNGVAVR 614
Qy 9 -----ELLV-----AVGPDVFOAHQEDTERY----- 29
Db 615 SNODLITENLLPGRELLQTNLINVTSIRDNIFVGRAEGTQYSKWYFVWVDEVTPFL 674
Qy 30 --VLTNINIG----- 37
Db 675 TAOATHLRVGVWALTGYTPYPGAGEGMGNGVGGDLYSGYDGLHLWTGHVAREVPTSPGQ 734
Qy 38 ----- 37
Db 735 HLLAPEDVISCLDLSPVSTISFRINGCPVQGVFESFNLDGLFPFVVSAGVKVRFLLGG 794
Qy 38 ----- 37
Db 795 RHGEFKFLPPGYAPCHEAVLPRERLHLEPIKEYRREGPRGPHLVGPSRCLSHTFVPCP 854
Qy 38 ----- 37
Db 855 VDTQIVLPHPLERIREKLAENIHELWALTREIQGWTVGPRVDDNKRLLHPCLVDFHSLPE 914
Qy 38 ----- 37
Db 915 PERNYNLQMSGETLKTLLALCHVGMADKAEADNLKTKLPKTYMNSNGYKAPAPLDLSHV 974
Qy 38 ----- 37
Db 975 RLTPAQTTLVDRLAENGHNWWARDVGQGSYSYSAVDQIPARRNPRLVPRLLDEATKRSN 1034
Qy 38 ----- 37
Db 1035 RDSLQAVRTLLGYNIETPPDQPSQVENQSRCDVRIFRAEKSYTVQSGRWYFEFAV 1094
Qy 38 -----AELLRDPISLCA----- 48
Db 1095 TTGEMRVGWARPELRPDVELGADELAYVFNCHRGORHLGSEPGRPWPQPDVVGCMIDL 1154
Qy 49 ----- 48
Db 1155 TENTIIFTNGEVLMSDSGSETAPREIEIGDGLPVCGLPGQVGHNLNGQDVSSLRFFA 1214
Qy 49 ----- 48
Db 1215 ICGLOEGFEPFAINMQRPVTTWFSKGLPQFEPVPLEHPHYEVSVDGTVDTPPCLRLTHR 1274
Qy 49 -----OFRVHL----- 54
Db 1275 TWGSONSLVEMFLRLSLPVOFHQFRCTAGATPLAPGLOPPADEARAAEPDPDYENL 1334
Qy 55 -----VK 56
Db 1335 RRSAGGWEAENGKEGTAKGAPGTPQAGGEAQPARAENEKDATTENKKNRGRGLFKAKK 1394
Qy 57 MVILTEPEGAPNI----- 69
Db 1395 VAMMTQPPATPTLPRPHDVVPADNRDDPEIILNTTTYYSVRVRFAGQEPSCVMAGWVTP 1454
Qy 70 ----- 69
Db 1455 DYHQHMSFDSLKVRVTVTWGDEQGNVHSSLKSCNCYMWVGGDPVSPGQQRISHTDLV 1514
Qy 70 -----TAN----- 72
Db 1515 IGCLVDLATGLMTFTTANGKESNTFFQVEPNTKLPFAVFLPTHQNVIOFELGKQKNIMPL 1574
```

```
Qy 73 ----- 72
Db 1575 SAAMFOSERKNPAPQCPPRLEQMQLMPVSWSRMPNHFLQVETPRAGERLGMVOCOEPLT 1634
Qy 73 ----- 72
Db 1635 MMALHIPEENRCMDILELSERLDLQRFHSHTLRLYRAVCALGNRVHAHALCSHVDOAQLL 1694
Qy 73 ----- 72
Db 1695 HALEDAHLPGPLRAGYYDLLISIHLESACRSRSMISEVIVLPTPETRAITLPPGRSTE 1754
Qy 73 -----LTSS----- 76
Db 1755 NGHPRHGLPGVGVTTSLRPHHPSPCFVAALPAAGAAEAPARLSPAIPEALRDXALRM 1814
Qy 77 ----- 76
Db 1815 LGEAVRGGOHARDPVGASVEFOFVPLVLVSTLLVMGIFGDEDEVKQILKMIPEVFTEE 1874
Qy 77 ----- 76
Db 1875 EBEDEBEDEBEDEBEDEBEDEBETAQEKEDKEBEDEBEDEBEDEBEDEBEGLLQMKLPES 1934
Qy 77 ----- 76
Db 1935 VKLQCHLLEFYCQOELQHRVESLAAFAERYVDKLOANQORSRYGLLIKAFSMTAAETARR 1994
Qy 77 -----LLSVCG----- 82
Db 1995 TREFRSPQEQINMLLOFQDGTDEDCPLPEEIRQDLDLDFHODLLAHACGIGLDEBEPE 2054
Qy 83 -----WSO----- 85
Db 2055 EETLGSRLMSLLEKVLVKKKEEPEERSAEESKPSRLOELVSHMVVRAOEDFVQSP 2114
Qy 86 -----TINPE--DDT----- 93
Db 2115 ELVRAMPFLLHROYDGLGELLRALPRAYTISPSVVEDTWSLLECLQIRSLIVQMGPQE 2174
Qy 94 ----- 93
Db 2175 ENLMIQSIGNMKNKVYQHPNLMRALGMHETVMEVMVNVNVLGGESKEIRFPKMTVSCCR 2234
Qy 94 ----- 93
Db 2235 FLCYFCRISRQNSRQSMFDHLSYLLENSGIGLMQGSTPLDVAASVIDNNELALALQEQD 2294
Qy 94 ----- 93
Db 2295 LEKVSYLAGCGLOSCPMLVAKGYPDIGWNPCCGERYLDLRFPAVFVNGESVENANVVV 2354
Qy 94 -----DP----- 95
Db 2355 RLILRKPECFGPALRGEGSGLLAAITEAIRISEDPAKDGPGRIRRRRRRHFGHEEPPEEN 2414
Qy 96 -----GHA-----DLV----- 101
Db 2415 RVHLGHAIMSFYAALIDLLGRCAPEMHILIOAGKEALRIRAILRSLVPLEDLVGIISLPL 2474
Qy 102 ----- 101
Db 2475 QIPTLKGKDALVQPKMSASFVPDHPKASMWLFLORVYGIENQDFLLHVLVDVGLPDMRAAA 2534
Qy 102 ----- 101
Db 2535 SLDTATFTTTEMALAVNRVLCVAVLPLITKCAPLFACTEHRAMVDSMLHTVVRSLRGRS 2594
Qy 102 ----- 101
Db 2595 LTKAQRDVEDCLMSLCRYIRPSMLQHLRLRVDFVPIELNEFAKMPKLKLTNNHYERCWKY 2654
```

QY 102 -----LVYTR-----FDLEL-----PD- 113
DB 2655 YCLPTGWANFGVTSBEELHLTRKLFWGFDSLAHKYDPELYRMAMPCLCALAGALPPDY 2714
QY 114 -----CN----- 115
DB 2715 VDASYSSKAEKKATVDAEGNDFPRPVETLNVIIPEKLDGFINKFAYETHEKWAFDKIQNN 2774
QY 116 ----- 115
DB 2775 WSYGENIDEELKTHPMLRPYKTFSEKDKEIYRPPIKESLKAMIANEWETIEKAREGESEKT 2834
QY 116 ----- 115
DB 2835 EKKTRKISQSAQTYDPREGYNPQPDLSAVTSLRELOAQAENYHNTWGRKKQBEL 2894
QY 116 ----- 115
DB 2895 EAKGGTHPLLVYDPTLTAKEKARDREKAQELLKFLQNMGYAVTRGLKMDLDSSSIEKR 2954
QY 116 ----- 115
DB 2955 FAFGLQQLLRHMDISOEFIAHLEAVVSSGRVEKSPHEQEIKFPAKILLPLINQYFNHC 3014
QY 116 ----- 115
DB 3015 LYFLSTPAKVLGGSHSNKEXEMITSLFCKLAALVRHRVSLFGTDAPAVVNCILHARS 3074
QY 116 -----RQVRGVTO----- 123
DB 3075 LDARTVMKSGPEIVKAGLRSFFESADIEKMNVLRLGKVSQARTQVKGVGQNLTYTTV 3134
QY 124 ----- 123
DB 3135 ALLPVLTTLFOHIAHQFGDDVLDQVQSVYRTLCISYSLGTTKNTYVEKURPALGEC 3194
QY 124 -----LG 125
DB 3195 ARLAAAMPVAFLEPOLNEYNACSVYTKSPRERAILGLPNSVEEMCPDIPVLERLMADIG 3254
QY 126 G----- 126
DB 3255 GLAEGARYTEMPHVIEITLPMLCYSLPRWERGPEAPPALPAGAPPCTAVTSDHLNS 3314
QY 127 ----- 126
DB 3315 LLGNLRIIVNVLGIDEASMMKRLAVFAQPIVSRARPELLOSHFPTIGRLRKRAKVVS 3374
QY 127 ----- 126
DB 3375 EEEQLRLEAKAEQEGELLVRDEFVLCRDLYALYPLLIRYVNNRAQWLTEPNPSABEL 3434
QY 127 ----- 126
DB 3435 FRMVGEIFIYWSKSHNFKREEQNFVVQNEINNMFLTADNKSMAKAGDIOSGGSDQERT 3494
QY 127 -----ACSP----- 131
DB 3495 KKKRGDRYSVQTSIVATLKKMLPIGLNMCAPTDQDLITLAKTRYALKOTDEEVREFLH 3554
QY 132 ----- 131
DB 3555 NNHLQGVKESPLRWQALYRGVPGREEDADDPEKIVRRVQEVSAVLYLDQTEHPYK 3614
QY 132 -----W-----LITEDTGF----- 143
DB 3615 SKKAVWHKLLSKORRRAVACFRMTPLYNLPTHRCNMFLESYKAAWILTEDHSFEDRW 3674
QY 144 ----- 143
DB 3675 DDLSKAGEQEEEEVEEKPPDLHLQLVLFHSRTALTEKSLDEDLYMAYADIMAKSCH 3734
QY 144 ----- 143

DB 3735 LEEGSEGEABEEVEVSFEKQMEKQRLLYQOARLHTRGAABMVLQMIASACKGETCAMVS 3794
QY 144 ---LGVTI 148
DB 3795 STLKLGISI 3803

RESULT 11

TYCC_BACBR
ID TYCC_BACBR STANDARD; PRT; 6486 AA.
AC O30409;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrocidine synthetase III [Includes: ATP-dependent asparagine
adenylase (AsnA) (Asparagine activase); ATP-dependent glutamine
adenylase (GlnA) (Glutamine activase); ATP-dependent tyrosine
adenylase (TyrA) (Tyrosine activase); ATP-dependent valine adenylase
(ValA) (Valine activase); ATP-dependent ornithine adenylase (OrnA)
(Ornithine activase); ATP-dependent leucine adenylase (LeuA) (Leucine
activase)].
GN TYCC
OS Bacillus brevis.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX NCBI_TaxID=1393;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 8185;
RX MEDLINE=98012987; PubMed=9352938;
RA Mootz H.D., Marahiel M.A.;
RT "The tyrocidine biosynthesis operon of Bacillus brevis: Complete
nucleotide sequence and biochemical characterization of functional
internal adenylation domains".
RT J. Bacteriol. 179:6843-6850(1997).
CC -1- FUNCTION: INCORPORATES SIX AMINO ACIDS (FOR TYROCIDINE A, ASN,
GLN, TYR, VAL, ORN, AND LEU) IN THEIR L-CONFIGURATION INTO THE
PEPTIDE PRODUCT.
CC -1- COPACTOR: CONTAINS 6 COVALENTLY BOUND PHOSPHOPANTHETHEINES (BY
SIMILARITY)
CC -1- PATHWAY: Cyclic peptide antibiotic tyrocidine biosynthesis.
CC -1- SUBUNIT: LARGE MULTIENZYME COMPLEX OF TYCA, TYCB AND TYCC.
CC -1- DOMAIN: CONSISTS OF SIX MODULES, AND HARBORS A PUTATIVE
THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE
INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
ADENYLATION, THIOYLATION, CONDENSATION (NOT FOR THE INITIATION
MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
(OPTIONAL).
CC -1- MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC
DECAPEPTIDES, TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-
ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR
RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
FAMILY.
CC -1- SIMILARITY: CONTAINS 6 ACYL CARRIER DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; AF004835; AAC45930.1; -.
DR HSSP; P14687; 1AMU
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Pantane_attach.
DR InterPro; IPR000379; Ser_estra site.
DR InterPro; IPR001031; Thioesterase.

DR Pfam; PF00501; AMP-binding; 6.
 DR Pfam; PF00550; DP-binding; 6.
 DR Pfam; PF00668; Condensation; 6.
 DR Pfam; PF00975; Thioesterase; 1.
 DR PRINTS; PRO0154; AMPBINDING.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
 DR PROSITE; PS00455; AMP BINDING; 6.
 DR PROSITE; PS50075; ACP_DOMAIN; 6.
 KW Ligase; Antibiotic biosynthesis; Phosphopantetheine;
 KW Multifunctional enzyme; Repeat.
 FT REPEAT 466 1038 DOMAIN 1 (ASPARAGINE-ACTIVATING).
 FT REPEAT 1521 2070 DOMAIN 2 (GLUTAMINE-ACTIVATING).
 FT REPEAT 2536 3113 DOMAIN 3 (TYROSINE-ACTIVATING).
 FT REPEAT 3590 4149 DOMAIN 4 (VALINE-ACTIVATING).
 FT REPEAT 4606 5203 DOMAIN 5 (ORNITHINE-ACTIVATING).
 FT REPEAT 5658 6245 DOMAIN 6 (LEUCINE-ACTIVATING).
 FT DOMAIN 970 1037 ACYL CARRIER (ACP) 1.
 FT DOMAIN 2007 2074 ACYL CARRIER (ACP) 2.
 FT DOMAIN 3045 3112 ACYL CARRIER (ACP) 3.
 FT DOMAIN 4080 4147 ACYL CARRIER (ACP) 4.
 FT DOMAIN 5124 5191 ACYL CARRIER (ACP) 5.
 FT DOMAIN 6167 6234 ACYL CARRIER (ACP) 6.
 FT BINDING 1000 1000 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 2037 2037 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 3075 3075 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 4110 4110 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 5154 5154 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 6197 6197 PHOSPHOPANTETHEINE (BY SIMILARITY).
 SQ SEQUENCE 6486 AA; 724011 MW; 4934900AF0DF786 CRC64;

Query Match 27 28; Score 210; DB 1; Length 6486;
 Best Local Similarity 1.3%; Pred. No. 1.3e+02;
 Matches 78; Conservative 32; Mismatches 33; Indels 6044; Gaps 23;

QY 3 GGILH----- 7
 DB 152 GVLLHLLTYAALKGEPIPREATKPYSEYIKWLDKQNKDEALAYQWYLAGYDHAAP 211
 QY 8 ----- 7
 DB 212 PKKLGTEASRYEHVEAMFTIAPEKTQOLIQAQNAQTSSVFQALWGILASTYKNADD 271
 QY 8 ----- 7
 DB 272 VVFGSVSGRRPQIQIESVMVGLFINTPTRVQTNKQOTFSELLQTVQKQALASATYDFA 331
 QY 8 ----- 7
 DB 332 PLYEIQTTVLKOELIDHLVTFENYPDHSMKHEESLGQFTVESGDEQTSYDLNVVVAL 391
 QY 8 ----- 7
 DB 392 APSNELYVKSNAAYSESVNRIEGHLRTVIDQVIGNPHVHLHEIGIITEEBEQQLLV 451
 QY 8 ----- 7
 DB 452 AYNDTAAEYPRDKTIFELIAEQASRTPAKAAVCGEDTLTYQELMERSAQLANALREKGI 511
 QY 8 -----LELVAV----- 14
 DB 512 ASGSIVSIMAEHSLELIVAMVLRSGAAYLPIDPEYQDRIQYLLDDSDQTLTLLTOSHL 571
 QY 15 ----- 14
 DB 572 QPNIRFAGSVLYLDDRLSYEGGSTFAPESKPDLDLAYMIYVTSGTGNPKGAMITHQGLVN 631
 QY 15 ----- 14
 DB 632 YIWANKVYQGEAVDPFLYSISFDLTVTSTIFTPLLSGNTIHVYRGADKVQVILDIKD 691
 QY 15 ----- 14
 DB 692 NKVGIILKTLPHLKLIEHIDGKASSIRRFVGGENLPTKLAKQIYDHGENVQIFNEYGP 751

QY 15 ----- 14
 DB 752 TETVGCMIYLYDPQTTOESVPFGVPADNVQLYLLDASMPQVPVPSGLGEMYIAGDGVAK 811
 QY 15 ----- 14
 DB 812 GYFNRPBLTKKEFIDNPFPRGTYMYRTGDLAKWLPDGNMEYAGRMDOYQKIRGHRIEMGE 871
 QY 15 ----- 14
 DB 872 IETRLTQHEAVKEAVIVKEKDEQNVLYAYLVSERELTVAELEREFLGRTLPSWYIPSF 931
 QY 15 ----- 14
 DB 932 IRLAEIPTANGVERKKLPKAGAVVTGTAYAAQNEIEAKLAEIMQVVLGISQVGIHD 991
 QY 15 ----- 14
 DB 992 DFFDLGGHSUKAMTVVQVSKALEVELPVKALFEHPTVAELRFLSRSEKTEYTAIQPVA 1051
 QY 15 ----- 14
 DB 1052 AQEFYPVSSAQKMYILQOPEGNGISYNISGAILLEGKLDYARFASAVQQLAERHEALRT 1111
 QY 15 ----- 14
 DB 1112 SFHRIDGEPVQVHEEVEVPLFMLEAPEQAQKIMREFVRPFDGLGVAPLMRTGLLLGKD 1171
 QY 15 ----- 14
 DB 1172 RHLFLDMHHIISGVSSQILLREFAELYOGADLOPLSLQYKDYKQFAAQNELFOTEAYKKQ 1231
 QY 15 ----- 14
 DB 1232 EQHWNLTFADEIPLNLPTDYPSPVQSPAGDLVLPFAAGKELLERLQOVASETGTTLMI 1291
 QY 15 ----- 14
 DB 1292 LLAAYNVLLSKYTGQEDIIIVGTPVAGRSHADVENIMGIFVNTLALRNQPASSKTFQFLO 1351
 QY 15 ----- 14
 DB 1352 EVKQNALAAADHODYPFEBELVEKLAIORISRNPLFDTLFSLENANQOOLAIABLTASPY 1411
 QY 15 ----- 14
 DB 1412 ELFNKISKFDLALNASESPADIQFLTATKLPFKKETVERMARHYLEILRWISEOPTASL 1471
 QY 15 ----- 14
 DB 1472 ADIDMMTEAKRTLNLNVNDTFVERTAATLHOLVBEQAARTPDEVAVVVEYALTREL 1531.
 QY 15 -----GPDV----- 18
 DB 1532 NARANQALRLRSHGTGPDTLIGIMYDRSPGMVGLAVLKAGAYTPIDPSYPERIQY 1591
 QY 19 ----- 18
 DB 1592 MLSDSOAPILLTQRHLQELAAVQGEIIDVEEAIYTGADTNLDNVAGKODLAYIVYVTS 1651
 QY 19 ----- 18
 DB 1652 TGNPKGWMISHQAICNHNMLWMRETPLTTEDAVLQKTPFSFDASVMEFVPLITGGOLV 1711
 QY 19 ----- 18
 DB 1712 AKPDGHRDIAYMRLIRDEKITTLOWVPSLLDWMTPDQWGSACTSLQRFVCGGEALTPAL 1771
 QY 19 ----- 18
 DB 1772 VSRFYETQQAQLINLYGPTTETIDATYWPQPRQOEYSATPICKPIDNVRLYVNVNASNQLQ 1831

Qy	19	-----	18
Db	1832	PVGAGELCIAGDGLARGYQWREBELTKASFDVDPFPGTMYRTGDMVRYLPDGHIEYLG	1891
Qy	19	-----	18
Db	1892	RIDHOVKIRGHRIELGEIEATLLOHEAVKAVVMARQDGKQNSLYAYVVAQODIQTAEL	1951
Qy	19	-----	18
Db	1952	RTYLSATLPAVMVPSAFVFLEQLPLSANGKVDKALPQPEDAAASAAVYVAPRNEWAKL	2011
Qy	19	-----	18
Db	2012	AA1WESVLGVEPICGHDHFFELGSHLKAMHWISLLQRSFQVDVPLKVLFPESPTIAGLAP	2071
Qy	19	-----	18
Db	2072	LVAARKGTYTAPPEVEKQYYPVPSAAQKRMFILOMEGAGISYNNPFGMYLQDKLQDTER	2131
Qy	19	-----	18
Db	2132	LQQALKSVLQVHESLRTSPHSVOGETVQVRHDDVDLAIISFGEATEAQIAEQFIQPPD	2191
Qy	19	-----	18
Db	2192	LGTAPLLRAGLIKLAPERHFLMDLHHIIVDGVSGILLIEEPAQLYHGEELPALURIQYKD	2251
Qy	19	-----	18
Db	2252	FAKQODWFOTEFABQEAQYMLNTFTGEIPLVNLPTDYPSPSVKSFAGDRFVFGTALP	2311
Qy	19	-----	18
Db	2312	KQLHQAQETGTTLYMVULLAANYVLLSKYSROEDIIVGAPTAGRSHAETESIVGMFNTL	2371
Qy	19	-----	18
Db	2372	ALRNEPAGGKTRDFLAEVKINTLGAPEHODYPLDELVDKLDMDQLSRNPLPDTVFILQ	2431
Qy	19	-----	27
Db	2432	NMEQKPEMEQLTITPYSAEVKQAKFDLSLEAYEENABIIIFSLDYSTKLSFRETIEKIAT	2491
Qy	28	-----	27
Db	2492	HFIQILRAVIAEPEMPLSEITMLTEAEKORLLVDENGAAHKDFPQNKTLQALFEEQAEKSP	2551
Qy	28	-----	27
Db	2552	QATAVEISGQPSYQELNERANQLAATLRERGVPQDPQVPGIMANRSVEMVVGILLAILKAG	2611
Qy	28	-----	40
Db	2612	GAYVPIDPEYPERVAYMLTDCQARLVLTQKHGAKLGSVTAECLYLDDDESNGVGHRSN	2671
Qy	41	-----	40
Db	2672	LQPINTASDLAYIIYTSGTTGPKGMVMEHRCIIVNNVLWKAQYQMKVGDORSLLSUSFAP	2731
Qy	41	-----	44
Db	2732	DAFVLSFFTVPVLSGATVVALEDEAKDPVSLKKLIAASRCTLMTGVPSLFQAILECSTPA	2791
Qy	45	-----	49
Db	2792	DIRPLOTVTLGGKEITAQLVEKCKQLNPDLVIVNVEYGPTESSVVATQWRLAGPDAAITIG	2851
Qy	50	-----	49
Db	2852	RPIANTSLYIVNQYHQLOPIGVVGEICIGGRGLARGYWNKPALTEEFKVSHPFAAGERMY	2911
Qy	50	-----	49

Db	2912	KTGDLGKWLP	PDGTIEYIGRIDBQVKRGYRIEIGETIESALLAAEKULTAAVVVYVDOLQG	2971		
QY	50	-----	-----	49		
Db	2972	SALAAFTADEQLDVTKLASHLSKRLPSYMI	PAHFVQLDOLPLTPNGKVKDKALPKPEGK	3031		
QY	50	-----	-----	49		
Db	3032	PVTEAQVYPTNAVESKLAEI	WERVILGVSIGILDNFFQITGCHSLKAMAVAAQVHREYQV	3091		
QY	50	-----	-----	49		
Db	3092	ELPLKVLFAOPTIKALAAQVYATSGKETYP	PIEPALQOEYYPVSSAQKRMVYVLROFADTGT	3151		
QY	50	-----	-----	49		
Db	3152	VYNMPSALYIEGDLDRKFEAAIHGLVERHES	LRTSFHTVNGEPPQVORVHEHVELNVOYAE	3211		
QY	50	-----	-----	49		
Db	3212	VTEAQVEPTVESFVQAFDLTKAPLLRVGL	FLKAAKRHLFLDDHHIISDGSAGIIMEEF	3271		
QY	50	-----	-----	49		
Db	3272	SKLYRGEELPALSVHYKDPAMVQSEL	FQSDVYTEHENYWLNAFSGDIPVLNLPA	3331		
QY	50	-----	-----	49		
Db	3332	TQSPGDCVSFOADKALLDDLHLKLAQES	OSTLFWULLAAYNVLLAKYSGOEDIWGTPIA	3391		
QY	50	-----	-----	49		
Db	3392	GRSHADIENVLGMFVNTLALRNPVETKH	FQAFLEBVKQNTLQAYAHQDYPPEALVEKLD	3451		
QY	50	-----	-----	49		
Db	3452	IORDLSRNP	LFDTMFILQNLDOKAYELDKLEAPQAQNAK	FOLTLBAHEDETGIFHA	3511	
QY	50	-----	-----	49		
Db	3512	LVYSTKLFQRESIERMAGHFLQVLROQVAD	QATALREISLSEBERRIVTVDFNNTFAAY	3571		
QY	50	-----	-----	49		
Db	3572	PRDLTIQELFEQQA	AKTPEHAAVMDGQMLTYRELNEKANLAHVL	RQNGVKGESIVGLL	3631	
QY	50	-----	-----	49		
Db	3632	ADRSLEMITGIMILKAGAYLGLDPEHP	SERLAYMLEDGGVKVVLVQKHLPLVGEGLM	3691		
QY	50	-----	-----	49		
Db	3692	PIVLEESLRBEDCNP	AIVNGASLAYMYTSGTGPKGMVEHRNVTRL	VMHTNYVO	3751	
QY	50	-----	-----	49		
Db	3752	VRESDRMIQTGAIGF	DAMTFEIFGALLHGASLYLVSKDVLDAEKL	GDPLRTNOITMWL	3811	
QY	50	-----	-----	49		
Db	3812	TSPLFNQLSQDN	PAMFDSLRLIIVGGEALSPKHI	NRVKSALPDLEIWN	GYGPTENTTFT	3871
QY	50	-----	-----	49		
Db	3872	CYLIEQHFEEQIPIG	KPIANSTAYIVDGNNOQPIGVP	PELCEVGGDVGARGV	VNKPDELTA	3931
QY	50	-----	-----	49		
Db	3932	EKFVNPFPAGETMYRTGD	LARWLPDGTIEYLGRIDQQVKIRGYRIE	ELGEITVLVSQAAQ	3991	
QY	50	-----	-----	49		

```
Db 3992 VKEAVVAIEEANGOKALCAFYVPEQAVDAAELEAMSKQLPGYMWVPAYVQMEKPLTA 4051
Qy 50 ----- 49
Db 4052 NGKVDRRALPOPSGERTTGSFAVAQNDTEAKLOQIWOEVILGIPAIGHDNFFBIGHSL 4111
Qy 50 ----- 49
Db 4112 KAMNVITQVHKTFQVELPLKALFATPTIHLEAAHIAESAPEQFETIQPVEPAAPYVSFA 4171
Qy 50 ----- 49
Db 4172 QKRMVILHQFEGSGISYNVPSVLVEGKLDYDRFAAAIQSLVKRHSIRTSFHSVNGEPL 4231
Qy 50 ----- 56
Db 4232 ORVHPDVLPVRLLEATEDQESLIQELIOPDLEIAPLFRVNLIKLGAERHLFFMDMH 4291
Qy 57 ----- 56
Db 4292 IISGVSLAVIEIASLYAGKQLSDLRIOYKDFAVWQTKLAQSDRFQKQEDFWTRTFAG 4351
Qy 57 ----- 61
Db 4352 EIPLLNLPHDYPRPSVQSGDGTVALGTGHHLEQLKLAETGTTLFWVLLAAYHVLLS 4411
Qy 62 ----- 61
Db 4412 KYAQEEIVGTPIAGRSHADVERIVGMFVNTLAKNTAAGSLSFRAFLEDVKQNALHAF 4471
Qy 62 ----- 61
Db 4472 EHODYPEHLVEKLOVRDLRSNPLFDTMFSGLAESAEGEVAOLKVPVPVNGHIAKFD 4531
Qy 62 ----- 61
Db 4532 LSLDAMEKODGLLVQFSYCTKLFKETVDRLAHVHVVQLLOITADPDIELARISVLSKAE 4591
Qy 62 ----- 61
Db 4592 TEHMLSHFLATKATYPTDKTFOKLFEQVEKTPNEIAVLFQNEQLTYOELNAKANQLARV 4651
Qy 62 ----- 61
Db 4652 LRRKGVPKSTVGILVDRSLYVMGLAVLKAGTVPIDPDYPLERQAFMLEDESEAKLL 4711
Qy 62 ----- 66
Db 4712 LTLOKMSOAVFPYETFYLDTETVDQEBTGHNLEHVAQBNVAYIITSGTTGPKGVVIE 4771
Qy 67 ----- 66
Db 4772 HRSYANVAFANKDEYHLDSFVRLQWASFAFDVSTGDFARALLTGQGLVCPNGVKRMDP 4831
Qy 67 ----- 66
Db 4832 ASLYETIRRHEITIFEATPALIMPLMHVYVENELDSOMKLLILGADSCPAEDFKTLAR 4891
Qy 67 ----- 70
Db 4892 FGQKRIINSYGVTEACIDTSYBETDVTAIRSGTVPFGKPLPNMTWYVVDHNLQPVG 4951
Qy 71 ----- 70
Db 4952 VVGELCIGGAGVARGYLNRPBELTBEKFPNPFAPGERLYRTGDLAKWRADGNVEFLGRND 5011
Qy 71 ----- 72
Db 5012 HQVKIRGVRIELGIETQLRKLDGITEAVVVARERDQEKELCAYVVDHKLDTABELAN 5071
Qy 73 LTSSL-----LSVCG----- 82
Db 5072 LKELPQNMIPAYFVTLDALPLTANGKVDRRSLPAPDVTMLRTTEYVAPRSVWEARLAQV 5131
```

```
Qy 83 WSQTIN----- 88
Db 5132 WEQVLNVPOVGALDDFFALGGHSLRAMRVLSSMHNEYQVDIPRILRILPEKPTIOELAAPIE 5191
Qy 89 ----- 88
Db 5192 ETAKGNVFSIEPVQKQAYYPVSSAOKRMVILDOFEGVGISYNNPSTMLIEGKLERTRVEA 5251
Qy 89 ----- 88
Db 5252 AFORLIARHESLRTSAFVNVGEPVQNIHEDVPFALAYSEVTEOEARELVSLSVQFPDLEV 5311
Qy 89 ----- 88
Db 5312 APLIRVSLKIGEDRYVLFTDMHHSISDVSSGILLAEWVOLYQGDVLPBLRIQYKDFAV 5371
Qy 89 ----- 95
Db 5372 WQOEFSSAAAFHKQEAAYWLOTFADDIPVLNLPTDTPTRPSTOSFAGDOCTIGACKALTEGL 5431
Qy 96 ----- 100
Db 5432 HOLAQATGTTLYMVLLAAYNVLLAKYAGOEIIIVGTPIITCRSHADLEPIVGMFVNTLAMR 5491
Qy 101 ----- 100
Db 5492 NKQREKTFSEFLOEVQKQNALDAYGHQDYPFEBELVEKLAIRDLSNPLPDTVTFQNST 5551
Qy 101 ----- 107
Db 5552 EEVMTLPBECTLAPFMTDETGOHAKFDLTTSATEEREEMTIGVYSLSLFTRETMEFSRH 5611
Qy 108 ----- 117
Db 5612 FLTIAASIVQNPHIRLGEIDMLLPEEKQIILAGFNDTAVSYALDKTLHQLFEEQVOKTPD 5671
Qy 118 ----- 126
Db 5672 QAALLFSEOSLTVSELNERANRLARVLRAKGVDPRLVAIMAERSPEMWIGILGILKAGG 5731
Qy 127 A----- 127
Db 5732 AYVPDPGYPQERIQYLLLEDSDNALLLSQAHLLPILLAQVSSSELPECLDLNAELDAGLSGS 5791
Qy 128 ----- 127
Db 5792 NLPVAVNQPTDLAYVIYTSGTTGPKGVMIPHOGIVNCLQWRDEYGGPSDKALQVPSFA 5851
Qy 128 ----- 129
Db 5852 FDGFVASLPAPLLGGATCVLPQBAAKDPVALKKLMAATEVTHYGVPSLQFQAILDCSTT 5911
Qy 130 ----- 137
Db 5912 TDFNQRCVTLGGEKLPVQLVQKTEKHPAIEINNEYGTENSIVVTISRSIAGQAITI 5971
Qy 138 ----- 137
Db 5972 GRPLANVQVYIVDEQHHLOPIGVVGBELCIGGAGLARGYLNKPBELTAEKFPANPRPOERM 6031
Qy 138 ----- 137
Db 6032 YKTGDLVKWRTDGTIEYIGRADEQVKRGYRIEIGIESAVLAYOGIDQAVVVVARDDDAT 6091
Qy 138 ----- 137
Db 6092 AGSYLCAYFVAATAVSVSGLRSHLAKELPAYMIPSYFVELDQPLSANGKVDRAKLPKPQ 6151
Qy 138 ----- 137
Db 6152 QSDATTREYVAPRNATEQQLAAIWQEVLGVEPIGITDQFFELGGHSLKATLLIAKVVEYM 6211
```

QY 138 ----- 137
Db 6212 QIELPLNLFQYPTIEKVADFIHKRFESRYGTAILLNOETARNVFCFTPIGAQSVYQK 6271
QY 138 -----EDT 140
Db 6272 LAAETOGVSLYSFDFIQQDNRMEOYIAAITADPSGPTLMGYSGGNLAFEVAKELEER 6331
QY 141 GFDLGV 147
Db 6332 GV--GV 6336
RESULT 12
ID ACVS NOCLA STANDARD; * PRT; 3649 AA.
AC P27743.
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Delta-(L-alpha-aminoacidipyl)-L-cysteiny-D-valine synthetase
DE (EC 6.-.-.-) (ACV synthetase) (ACVS).
GN PCBAB.
OS Nocardia lactamurans.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;
OC Amycolatopsis.
OX NCBI_TaxId=1913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VAR LC 411;
RA MEDLINE=92065808; PubMed=1956290;
RX Coque J.J.R., Martin J.F., Calzada J.G., Liras P.;
RT "The cephamycin biosynthetic genes pcbAB, encoding a large
multidomain peptide synthetase, and pcbC of Nocardia lactamurans are
clustered together in an organization different from the same genes
in Acromonium chrysogenum and Penicillium chrysogenum.";
RL Mol. Microbiol. 5:1125-1133(1991).
CC -!- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE
ACV ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS
FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER
INTERMEDIATES.
CC -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETHEINES.
CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND
CEPHALOSPORIN.
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
FAMILY.
CC -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57310; CAB40561.1; -
DR PIR; S18268; S18268.
DR HSP; P14687; LAMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Ppantne attach.
DR InterPro; IPR000379; Ser extra site.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00501; AMP-binding; 3.
DR Pfam; PF00550; pp-binding; 3.
DR Pfam; PF00668; Condensation; 3.
DR Pfam; PF00975; Thioesterase; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.
DR PROSITE; PS00455; AMP BINDING; 1.
DR PROSITE; PS00075; ACP_DOMAIN; 3.

KW Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
KW Repeat; Phosphopantetheine.
FT REPEAT 401 861 DOMAIN 1 (ADIPATE-ACTIVATING).
FT REPEAT 1014 1937 DOMAIN 2 (CYSTEINE-ACTIVATING).
FT REPEAT 2079 2985 DOMAIN 3 (VALINE-ACTIVATING).
FT DOMAIN 788 857 ACYL CARRIER (ACP) 1.
FT DOMAIN 1864 1933 ACYL CARRIER (ACP) 2.
FT DOMAIN 2910 2981 ACYL CARRIER (ACP) 3.
FT BINDING 820 820 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 1896 1896 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 2944 2944 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT ACT SITE 3502 3502 THIOESTERASE (BY SIMILARITY).
SQ SEQUENCE 3649 AA; 404079 MW; 6FD095704F85986B CRC64;
Query Match 27.0%; Score 209; DB 1; Length 3649;
Best Local Similarity 2.6%; Pred. No. 27;
Matches 81; Conservative 23; Mismatches 43; Indels 2965; Gaps 24;
QY 1 AAG-----GIL----- 6
Db 180 AAGCLHRIAYAGEFFDKIIAGVLDVAREVLGQFGRPEQLVADIDLVSABOELQHQW 239
QY 7 ----- 6
Db 240 NCTGEPEDKRLNELFEDVVRAPDREAVVCCDVRLTYREVNERNQFAHMLIOGPVRV 299
QY 7 -----H 7
Db 300 RPGALIGLYDKSLGVVATFGIMKSGAAYVPIDPAYPAERIRFLVGTGLSGIVTNRH 359
QY 8 LELL-----VAVGPD----- 17
Db 360 AERLEVLGDEHSAHVHVEAVVAGPHPEQARENPGALSRRDAYVYVYTSGGTGPVK 419
QY 18 -----VFOAHOEDT----- 26
Db 420 VPKHYSVNSITDLSERYDMRRPOTERVALFASVVFPHLRLQTLIALINEQTLVIVPD 479
QY 27 -----ERYVLTNLN----- 35
Db 480 VRLDPDLFPEYIERHGVTYLNATSGVLQFHLRRCASLKLVLVGEELTASGLRQREKF 539
QY 36 -----TGA----- 38
Db 540 AGRVVEYAFTEAFVTAKEFGVGTERRDRSIGRPLRVNWKVYLSQGLKOLPIGAIGE 599
QY 39 ----- 38
Db 600 LYIGCGVAPGLNRRDDLTAERFTANPQTTEEKARGNRLYRTGDLARVLLNGEVEPM 659
QY 39 ----- 38
Db 660 GRADFQKLVGRVPEGEIAQAQTEFPVGVKCVVAKENATGDRHLVGVYLVEDGAETAE 719
QY 39 ----- 38
Db 720 ADLIAFLQRLIRIMVARMVRLTSPVNVNGKVDWRALPDVSLHPAPANMNGALLAID 779
QY 39 ----- 38
Db 780 GSNAPLLAITEOLRAIWSEVLGVQPNRIGERDDPFRLLGQSISCILLIARVQRSLSLG 839
QY 39 ----- 38
Db 840 VEDFALRTLALAGHLESQGHAEPEVVAEEVTTGSEPVRLVLANGLOOGLLYHLLKTAGG 899
QY 39 -----ELLRD-----PSL----- 46
Db 900 DDAYVQSVHRYHAPIRPELMKDAQAAQRTYPAALRLRFDWAEPEVQIVDNDKPPDMRF 959
QY 47 -----GAQFRVHLVKM----- 57
Db 960 VDLSATADDAEQEARVRELQERDRTEPYDLAGGLRFLRVYLTKQREDLFLIFSCHHILD 1019

QY	58	-----	57	QY	93	-----	92
Db	1020	GWSLPVLHDEVHRNYLALRAGOPTESVDNAYVAAQRYWEAHRNDHAAVVEQLGRIDER	1079	Db	2100	DGYPEADLGLRGTYGEFAVMQORYLTGKRLAALTSEFWTGCALGPFETIALPDHPRPRF	2159
QY	58	-----	57	QY	93	-----	92
Db	1080	GDFAGLLNEKSRYSVLGDYDHVQHRTRKLYLGADLTGALKAGCAADQVTLHSLVQFVM	1139	Db	2160	DYRGRELEFELDERTTREALRELARTARVSLYSVLLGAWCLMLNMYTQGHDLVVGTPSANR	2219
QY	58	-----	57	QY	93	-----	104
Db	1140	HKVLHAIGGGNTTVGTIVSGRNLPGDIENSAGLINTPLIIVDHQQQAGQNVAAEVRD	1199	Db	2220	GRPEFORAVGFFPANLLALRVRVDPDA-ATLPAYVRSVGEAVVAAQVHGELFPEQLVKELKV	2278
QY	58	-----	57	QY	105	-----	110
Db	1200	IOAAVNTWNSKSVIELGROSGEMKRRLFDTLVLENYPRLLDEEBELAHQEARFEKAY	1259	Db	2279	EKDPSPRHPILQNLFTLQNVSDHTSALTGYQPDGGMTTKTFDLTSATMTETATGLAGNLTY	2338
QY	58	-----	57	QY	111	-----	116
Db	1260	DADKVDPYIAVAREEGDELTVTLWYAGELFDEDTIDTLDDVARTLFRQVTEDIARPVRE	1319	Db	2339	AASLFDSDTSASGFIATFKHVLAEFASAAAQTPIAQUTALDEPQOAAALPDATRRARRPGGP	2398
QY	58	-----	57	QY	117	-----	116
Db	1320	LDLISPMRARFDSMNETAEFPADKTLHAVFEEMAERWPDEIAVYRENRLTYRELNER	1379	Db	2399	GRCTRLEFEVAATWPDRAVAVHGDVRLTYRELNERANRLAHHLRSVAEPRADELIALVLD	2458
QY	58	-----	57	QY	117	-----	123
Db	1380	ANRLAHLRSVVELRPDDLVALVLDKSELMTAIIAAWKTGAAVVPIDSGYPDDRISEML	1439	Db	2459	KSELTILVAIIAVNKAGAAVMPIDPSYPDDRIAFLMSDTGAKVLAGEAHGSRVRLTSGD	2518
QY	58	-----	57	QY	124	-----	123
Db	1440	SDTAARVVVWNEIHSRDLRSIAETGTPVLEIETELLHDDQPAVNPVTETSTDLAYAYTS	1499	Db	2519	VLDLEQLDLTGEPAPNPVTETSTELAYAIYTSGTTGPKKAVLVSHSGVDSFRAQLSGRY	2578
QY	58	-----	73	QY	124	-----	126
Db	1500	CTTGKPKAVLVEHGRVNVQLSLAKLFGDKAHKDEALLSFSNYIFDHFEQMTDALLNG	1559	Db	2579	FGSPDESABEAVFLANYVDFSVBQLALSVLGGHKLVPSPSAADDPAFYELANREGLSY	2638
QY	74	-----	79	QY	127	-----	126
Db	1560	QKLVVLDSDMRDQRLCRYNDEQVTVLSGTPSVLSLYDYSSATSLTRIDAIGEDTEP	1619	Db	2639	LSGTPQVERFDLRLSHLRCVLVAGEAFQPOHFEMKRGFAGPILNAYGTTTETTVNTV	2698
QY	80	-----	79	QY	127	-----	126
Db	1620	VFAKIRCTFPGLIINGYPTESISITSHKRPYPDPVHRVKNKSGIPPVANTKCHVLNKAMKP	1679	Db	2699	HRPEPGDAYRNTLGAPLGNTRLVYLGDMKMLPTGAVGELYLAGDCVTEGYLHRPELTRE	2758
QY	80	-----	79	QY	127	-----	127
Db	1680	VPVGGIGELYIGGIVTRGYLNREDLTADRFVENPFOFAEBRRLGENGRLYKTGDLVRWL	1739	Db	2759	RFLPNPFAESGRFPMTIYRTGDIVVRGPDGELQYLGNDQAQVKINGLRIEPEGEVEAALAG	2818
QY	80	-----	79	QY	128	CSPTWSCL1	136
Db	1740	PNGEVEYLGRTDLQVKIRGORVELGEVEAALSSYPGVVRSLVVAREHAVGOKYLVGFVVG	1799	Db	2819	CSGVROCAVVAGADPOAPERKRLVGYVYLPPEGAADVDEADLFAALRAQLMPSMVPSLLVRL	2878
QY	80	-----	79	QY	137	-----	136
Db	1800	EQEFDEQDLKQWMRKLPESVVPARVLRTIDIPVTPSGKLDARLPEPDFGAGEGEYVA	1859	Db	2879	DRPLPMTITGKLDVADLPADSPKRAAYAAPDRVRLEARLCHLWSAQLPGTGVGIDDDFF	2938
QY	80	-----	88	QY	137	-----	136
Db	1860	PVSEFELKLCGIWAQVLEIAPDRIGVHDDFFALGSDSIRAMALAQAITTFGCOGLGVATV	1919	Db	2939	RCGGDSISALHLASQVOREIERKSVKYLFDHPTVRSFVDNVLISGLAESGDDEPEOGR	2998
QY	89	-----	88	QY	137	-----	136
Db	1920	LOHTTLAAQAEHIQAAALEHTAWPTTAVEHPVPSLAQERLLFFIDDFEGGTAAVNPVFV	1979	Db	2999	TGECPLPIQEWFFAKPLADHRHWNHFAIRTPPLDPGELRTALDRLVEHHDAPRLRPE	3058
QY	89	-----	92	QY	137	-----	136
Db	1980	LRLPAHTRALPGALGTLVRRHPALRTLKTDQGRQYPIPADDDVRLVPSITVDSRA	2039	Db	3059	SGGEVYAEADAAPITLHELDVRLGADADLRLVDMORTFDLANGPTCAAYLHGHFDDGTA	3118
QY	93	-----	92	QY	137	-----	136
Db	2040	ELDEVLTIRAGVYVFRLHBEELPIRAEAFDHGDEIYLSVVVHHSCFDGWSWDIFRRELAALL	2099	Db	3119	RWPFALHHLVVDTVSWHILAOQLEILYNGGDLGNKTSYRQWAAQVRDYPAEGEREFWA	3178
				QY	137	--TED--	142

Db 3179 ETTROMESAELLAQTGTTTTRREBFALTAPDTRTLLAESPMAYDTEVNDLLLTATGFAIR 3238
 QY 143 -----DLGVTI 148
 Db 3239 SITROATNHLTVEGHGRELFCGAPDVRTVGVFTTWHFPAVEVDPGLGRSV 3290
 RESULT 13
 ID FAT_DROME STANDARD; PRT; 5147 AA.
 AC P33450; Q9VQX5;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cadherin-related tumor suppressor precursor (Fat protein).
 GN FT OR CG3352.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92069752; PubMed=1959133;
 RA Mahoney P.A., Weber U., Onofrechuk P., Biessmann H., Bryant P.J.,
 RA Goodman C.S.;
 RA "The fat tumor suppressor gene in Drosophila encodes a novel member
 of the cadherin gene superfamily.";
 RL Cell 67:853-868(1991).
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN. ACTS AS A

CC CC TUMOR SUPPRESSOR. REQUIRED FOR CORRECT MORPHOGENESIS.
 CC -1- SIMILARITY: CONTAINS 34 CADHERIN DOMAINS.
 CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M80537; AAA28530.1; --
 CC EMBL: AE003577; AAF51036.1; --
 CC PIR: A41087; IJFFTM.
 CC HSSP: P00740; 1EDM.
 CC FlyBase: FBgn001075; ft.
 CC InterPro: IPR002126; Cadherin.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR000742; EGF 2.
 CC InterPro: IPR001791; Laminin_G.
 CC Pfam: PF00008; EGF; 4.
 CC Pfam: PF00028; cadherin; 34.
 CC Pfam: PF00054; laminin_G; 2.
 CC PRINTS: PR00205; CADHERIN.
 CC SMART: SM00112; CA; 34.
 CC SMART: SM00181; EGF; 4.
 CC SMART: SM00282; LamG; 2.
 CC PROSITE: PS00232; CADHERIN_1; 22.
 CC PROSITE: PS02068; CADHERIN_2; 34.
 CC PROSITE: PS00022; EGF 1; 4.
 CC PROSITE: PS01186; EGF 2; 2.
 CC PROSITE: PS00025; Lam_G DOMAIN; 2.
 CC Cell adhesion. Signal; Transmembrane; Glycoprotein; Calcium-binding;
 CC Repeat; EGF-like domain.
 CC SIGNAL 1 35
 CC CHAIN 36 5147
 CC DOMAIN 36 4583
 CC TRANSMEM 4584 4609
 CC POTENTIAL.
 CC CYTOPLASMIC (POTENTIAL).
 CC CADHERIN 1.
 CC CADHERIN 2.
 CC CADHERIN 3.
 CC CADHERIN 4.
 CC CADHERIN 5.
 CC CADHERIN 6.
 CC CADHERIN 7.
 CC CADHERIN 8.
 CC CADHERIN 9.
 CC CADHERIN 10.
 CC CADHERIN 11.
 CC CADHERIN 12.
 CC CADHERIN 13.
 CC CADHERIN 14.
 CC CADHERIN 15.
 CC CADHERIN 16.
 CC CADHERIN 17.
 CC CADHERIN 18.
 CC CADHERIN 19.
 CC CADHERIN 20.
 CC CADHERIN 21.
 CC CADHERIN 22.
 CC CADHERIN 23.
 CC CADHERIN 24.
 CC CADHERIN 25.
 CC CADHERIN 26.
 CC CADHERIN 27.
 CC CADHERIN 28.
 CC CADHERIN 29.
 CC CADHERIN 30.
 CC CADHERIN 31.
 CC CADHERIN 32.
 CC CADHERIN 33.
 CC CADHERIN 34.
 CC CADHERIN 35.

FT	DOMAIN	3546	3651	CADHERIN 33.	QY	10	9
FT	DOMAIN	3652	3756	CADHERIN 34.	Db	258	VTILDVNDNPPIDFHSYDYNVSLNETALPGTPVTVTMASDNDLGDNSKITYYLAETEHQFT 317
FT	DOMAIN	3950	4011	EGF-LIKE 1.	QY	10	9
FT	DOMAIN	4013	4049	EGF-LIKE 2.	QY	10	9
FT	DOMAIN	4052	4090	EGF-LIKE 3.	Db	318	VPETGVISTTERVNCPCQQTNVKSSASOKSCVTFVFARDHGSPRODGRYTVTVNLLDTND 377
FT	DOMAIN	4092	4128	EGF-LIKE 4.	QY	10	9
FT	DOMAIN	4129	4320	LAMININ G-LIKE 1.	Db	378	HDPIISPRFPDGGKATVDENAVNGTVAAVAVKSDSGLNGRTSRVIVSGNELGHFRL 437
FT	DOMAIN	4321	4362	EGF-LIKE 5.	QY	10	9
FT	DOMAIN	4402	4569	LAMININ G-LIKE 2.	Db	438	EEAADLHVRVNGVLDREEIGKYNLTVMVMDQGTPTTTTAHLIIDVNDVNDHPEVPFKS 497
FT	DISULFID	3954	3966	BY SIMILARITY.	QY	10	9
FT	DISULFID	3960	3999	BY SIMILARITY.	Db	498	EYSNLSLAPGTSFVASITATDEDTGVNAQVHYDILSGNELKWFMDPLTGLIVTTGFL 557
FT	DISULFID	4001	4010	BY SIMILARITY.	QY	17	16
FT	DISULFID	4017	4028	BY SIMILARITY.	Db	558	DREIRDTVELSISARDGGPNPKPAYTLKVIILDENDEAPQFSOREQNVLTGEDAPPOTI 617
FT	DISULFID	4022	4037	BY SIMILARITY.	QY	17	16
FT	DISULFID	4039	4048	BY SIMILARITY.	Db	618	VALMTATDHDQGTNGSVTFALAPSVERLYPLQFALDAUTGOLTTTRRPLDREKMSQYEISV 677
FT	DISULFID	4056	4067	BY SIMILARITY.	QY	17	16
FT	DISULFID	4061	4078	BY SIMILARITY.	Db	678	IARDQAPTPOSATATVWLVNADVNDNDPQFYPRHYIYSLADDDDDIKLKEVEKERILL 737
FT	DISULFID	4080	4089	BY SIMILARITY.	QY	17	16
FT	DISULFID	4096	4107	BY SIMILARITY.	Db	738	HVTASDKDDGDNALIEYRLSEGGELFOLDARSGAISLRGAPASMHMKPHYKLLVSARD 797
FT	DISULFID	4101	4116	BY SIMILARITY.	QY	17	16
FT	DISULFID	4118	4127	BY SIMILARITY.	Db	798	AGORRQODAIVEIVLKSLEMBEGQAAGYBQMVNEDHEQQRNSQPNREVIGIVOKS 857
FT	DISULFID	4325	4341	BY SIMILARITY.	QY	17	16
FT	DISULFID	4334	4350	BY SIMILARITY.	Db	858	TNGKANSHEIYDIIOGDPAQNFRIIDTRSGRITARPPLDREBOANYRLTILASSSSSSAA 917
FT	DISULFID	4352	4361	BY SIMILARITY.	QY	17	16
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. .)	Db	918	ASSVSYGQCIYVIAIIDLNDNAPVFPALDRESEPTISLPENAAVGOEIVLSRVDRDAGVN 977
FT	CARBOHYD	257	257	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	276	276	N-LINKED (GLCNAC. .)	Db	978	SRISYSLTNPNQOFRIGPVTGVLYLQRPRAEPGSLIHVELMATDAGSPPLSSKLSLV 1037
FT	CARBOHYD	280	280	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	280	280	N-LINKED (GLCNAC. .)	Db	1038	LIADVNDHTVPDHTSYETSULPETTKVTRFPALAAATDIDLCDNGRISYEIIIEGNTERMF 1097
FT	CARBOHYD	402	402	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	461	461	N-LINKED (GLCNAC. .)	Db	1098	GVFPDGYLFVRAPLDRREEDYVYALTVSCRDAGQPSRSSVVPVVIHVIDENDNAPQPTNST 1157
FT	CARBOHYD	605	605	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	631	631	N-LINKED (GLCNAC. .)	Db	1158	FTFSPENAPADTFVGKLTAVDRDIGNAELFTLSSQOTDFTIDTRNGFIKTLRPFDR 1217
FT	CARBOHYD	1155	1155	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	1367	1367	N-LINKED (GLCNAC. .)	Db	1218	ALVKVSRNAESGDSGLRSGMAGNYMLLEATVSDNGIPRLQDKVKVKVITVDNDNAPE 1277
FT	CARBOHYD	1458	1458	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	1751	1751	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	1831	1831	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	1880	1880	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	2080	2080	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	2171	2171	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	2247	2247	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	2290	2290	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	2437	2437	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	2581	2581	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	2799	2799	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	2920	2920	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	2946	2946	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	2967	2967	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	3167	3167	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	3303	3303	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	3386	3386	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	3389	3389	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	3525	3525	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	3852	3852	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	3865	3865	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	3905	3905	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	4306	4306	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	4414	4414	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	4471	4471	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	4487	4487	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	4539	4539	N-LINKED (GLCNAC. .)	QY	17	16
FT	CARBOHYD	4550	4550	N-LINKED (GLCNAC. .)	Db	1278	FLRAPHVHTISEGASEGTHITHVFTQDADEGLNGDVVYSLAKGNEAQGFNLDSATGQLSL 1337
FT	CARBOHYD	1229	1229	S -> G.	QY	17	16
FT	VARIANT	1233	1233	G -> S.	QY	17	16

Query Match

Beet Local Similarity 1.8%; Pred. No. 76;

Matches 86; Conservative 23; Mismatches 36; Indels 4763; Gaps 29;

2 AGGI-----LHLE-----9

198 AGVNDKFLVTTANPSGDTSYLHLETTGNLDRSGSYQLNISARDGSPRFGYLQVN 257

```
Db 1338 GRRLDRESQEIHHILIVAKDAALKHPLSSNASITIVILDENDNAPFTQSSSEVSLETS 1397
Qy 17 -----16
Db 1398 PTGTELMFRASDADOGVNSQVVFISAGNRRDTHIDSITGSLYLHKPLDYEDITSYTL 1457
Qy 17 -----16
Db 1458 NITASDCGTPSLSTTVLYNVLDNDNPPFPSTAIVRQIKEGIPLKPTIVTVTADDPD 1517
Qy 17 -----16
Db 1518 SCLNGKSVYAISKQBPOLPOGRHFGINTETGVIHTLREIDRESIDTFLTVVATDRAQPS 1577
Qy 17 -----16
Db 1578 ERQSTKLVTVIVEDINDNAPVFSVMNAAILPPKFSTSGSSTAVMQVHAKDADSSNG 1637
Qy 17 -----16
Db 1638 LVTYIVSGPOELFKLQRTGIIITPGPQKQEVRYQLTKSTDEAVQSERRSSEVIT 1697
Qy 17 -----16
Db 1698 IITPGSGSESVPOFEQORSKLSGVYENEPIGTSILTVAHLASAEIEYFVTNVTATGS 1757
Qy 17 -----16
Db 1758 RGQVDRLPDIDAKGILSTAAELDREAGPEEYEVYAIAGLGQOPRTSRTKRVTVLDKN 1817
Qy 17 -----16
Db 1818 DSPQFLOTFPVYNVSEDLQIGHTISTLRAHPDPTLGSVTFLLMDGHGKFLLEPSTGKL 1877
Qy 17 -----16
Db 1878 ILNDTLDRETKSKYELRIRSDGOVTEAYATIOVSDTNDNPPLPFDTVYSPDIPENAAQR 1937
Qy 17 -----DVFO-----20
Db 1938 GYQVGQIVARDADLQNAQLSYGVVSDWANDVFSINPOTGMLTLTLARLDYEEVQHYILIV 1997
Qy 21 -----20
Db 1998 QAQDNGQPSLSTTITVYCNVLDLNDNAPIFDPMSYSSEVFENVPIATEVWTVSAKDIDSG 2057
Qy 21 -----20
Db 2058 NNGLIYSITAGDVDFGIDSNGTIRTRNLDREHRSYTLTVTARDCADEFASFSELE 2117
Qy 21 -----AHQ-----23
Db 2118 ETQLKLYRSRKYQOTROEFLAHQQRSLSTVKVITILIKDVNDVFPVFSANETAIME 2177
Qy 24 -----23
Db 2178 NVAINTVIAKAVDNDEGRNGYIDYLMKEARDEDMGQSDPLPFSINPTDGLRVVDALD 2237
Qy 24 -----23
Db 2238 RELRSSYLLNITARDGEPPQSTESQLLIRILDENDNSPVDPKQYSASVAENASIGAMV 2297
Qy 24 -----EDT-----ER-----28
Db 2298 LQVSATDVEGANGRIRYSIVLGQNHDFSISEDGTGVRVAKNLNYELSRYSLTVRAED 2357
Qy 29 -----YVLT-----32
Db 2358 CALENPAGDTAELTINILDINDNRPTFLDSPYLARMENTVFPNGGYVLTVNAYDADTPP 2417
Qy 33 -----32
Db 2418 LNSOVRYFLKEGSDSLFRINASSGDIALKPLDREQOSEYTLTLVAMDTGSPPLTGTGIV 2477
Qy 33 -----NLNIGAEEL-----RDPSLGAOFR-----51
Db 2478 RVEQDINDNDPVPFELQSYHATVRENLPSTGTHVLTPRATDKDEGLNAKRLNGLGEHMR 2537
Qy 52 -----VHLVKM-----57
Db 2538 FHIDSETGEISTATTLDRBEYSVHLTMAQDSSITEPRASSVNLTISVSDVNDONIPKPD 2597
Qy 58 -----57
Db 2598 STTYNAVPERISKGEFVGARALDLDGGENAVVHYTISGRDQHYFDINTKTGVVSTKLE 2657
Qy 58 -----VILTEPE-----64
Db 2658 LKTKTSHDDLTYTIVISAMDQEOSLSKAEALTILRPPELFPPTFAYMANSHFAMSBDV 2717
Qy 65 -----64
Db 2718 RPKXMITKVSATSPKGLVGKIRYAIAGGIMGDSLRVDPNSGLLSVGQDGLDYELTHLYE 2777
Qy 65 -----64
Db 2778 IWIEAADGDTPLSRSVTLITLNVTDANDNAPVMEQLIYNAEVLSEESPQLIAVVKASDR 2837
Qy 65 -----GAP-----67
Db 2838 DSGDNGNVIYRLQNDPDGTFEITESGEIYTRMLRDREBIGDYAFVVEAVDQGVPHMTGTA 2897
Qy 68 -----NITAN-----LTSS-----76
Db 2898 SVLHLLDKNDNPKPKFTRFLFSLNVTEAIEGFSVIRVTSOLDIGANANASYSFSENPE 2957
Qy 77 -----76
Db 2958 KFRIEPOSGNITVAGHLDRQODEYILKVNASDGAWEAETPITITIQDNDNAPEFHSF 3017
Qy 77 -----76
Db 3018 YSPFPPELQOSIALVGQIIATDRDKQGNPSVISYSLOQSPMFSIDPATGEVFSKAVRF 3077
Qy 77 -----LLSVC-----81
Db 3078 KHSQVRSPPENMYALTVLATDNGKPLYSECLVNIINIVDAHNNPKPEQAEYLAPLPDA 3137
Qy 82 -----GWSOTINP-----89
Db 3138 VRGQIRVVRHANDKQDLGTNEMDYSMTFNLSSIFSVRHGDGWLTVKPIQVPPNTRYEL 3197
Qy 90 -----BDD-----92
Db 3198 VVRATDRGVPQSDETRVIIVVTGENMDTPRFSVNSQVIVPENEPVSGTILTVGATDDD 3257
Qy 93 TDPG-----96
Db 3258 TGPNGMLRYSISGGNERQDFSVDBERTGGIVIQOQLDYDLIOEYHLNITVDGLGYHPLSSV 3317
Qy 97 -----96
Db 3318 AMLTILTDVNDNPPVFNHKEYHCYIPENKPVGTFFQAHAAKDPSKNAIHYAFLPSG 3377
Qy 97 -----96
Db 3378 PDRHFFIMQSNGTISSAVSFDYEBERRIYTLQIKAKNPSSSMESYANLYVHVLGVNFPYP 3437
Qy 97 -----HADL-----VLY 103
Db 3438 QFLOPVFHFVDFVSETSAVGTRVGAVQATDKSGEDGRVYLLVGSNDKGFRIIDTNTGLTY 3497
Qy 104 ITR-----106
Db 3498 VARHLDRQNRVVLTVNAKNYGSIRGNDTDEAQVVIISIQDGNDDPPEFIKHYTSTISEA 3557
```

```
QY 107 ----- 106
Db 3558 APVGTKVTTKAKDKVTONNOPSYSIINGNLKQSFKIDVQTGEISTASRLDRETSY 3617
QY 107 ----- 106
Db 3618 NLVIGAITDGLPQTGSATVHIELEDVNDNGPTFTPEGLNGYISENEPAGTSMITLIASD 3677
QY 107 ----- 111
Db 3678 PDLPRNGPPTYQLIGGKHKSWLSVDRNSGVVRSTTSFDREMTPILEAIIIEVEDSGKPKQ 3737
QY 112 ----- 111
Db 3738 KSOHLITITVLDQNDNPSTRSLHTIAVSLFNGDLPNSVKLADVRPNIDIVGDYRCRLQK 3797
QY 112 ----- 111
Db 3798 NPAQSQLQAI PRACDLITTSHTTPIASVSYTGNDGKHGDVSSKVSVAFOFNNETLAN 3857
QY 112 ----- 111
Db 3858 SVSIVMVRNMTAYHFLANHYRPILEMIKSRMSNEDEVILYSLLEGGSGNSTNLQLLMAVRL 3917
QY 112 ----- 111
Db 3918 AKTSYQPKYLIBRLREKRSFAFSELLQKEVIVGVEPCSEPDVCGNGVCSATMRLDDAHS 3977
QY 112 ----- 111
Db 3978 FVIQDSPALVLSGRVVRVHDYSOCCTSGFGSQCRSRQDPCLPNPCHSQVQCRRLGSDFCQ 4037
QY 112 ----- 114
Db 4038 MCPANRDGKHEKERSDVCYSKPCRNCGSCQSPDGSYFCLCRPGRGNQCESVSDSCR 4097
QY 115 ----- 114
Db 4098 PNPCLHGLCVSLKPGYKCNCTPRYGRHCFRFSYGFQPLSYMTFPALDVTNDSIVFA 4157
QY 115 ----- 114
Db 4158 TTKPNSLLLYNMQSGGRSDFLAIELVHORAYFSSGGARTALSTVIAGNLADGWHKV 4217
QY 115 ----- 122
Db 4218 TATRNGRVMSLSVAKCADSGDVCTECLPGDSSCYADEVGPVGTLPNFKNKQPLMIGGLSSAD 4277
QY 123 ----- 131
Db 4278 PILERPQVHSDDLVGLCHSVHIGGRALNLSPLQOKGILAGCNROACQALAAERCGGF 4337
QY 132 ----- 133
Db 4338 AGOCIDRWSSSLCQCGHLOSPDCSDSLEPITLCEGAFVFRISEYRRMQLLDNLNYSK 4397
QY 134 ----- 133
Db 4398 SANLDNQMRERRAVSNFSTASQIYEAPKMSMLFRITYKQGGQILYAATNQMTLSLRE 4457
QY 134 ----- 133
Db 4458 GRLVYYSKQHLINMTVQETSTLNDGKWHNVSLFSESRLIRLVDRQVGDDELDIAGVHD 4517
QY 134 ----- 136
Db 4518 FLDPYLTILNVGGEAFVGCCLANVTVNNELQPLNGSGSIFPEVRYHGKIESGCRDGTQDA 4577
QY 137 ----- 136
Db 4578 AQVADPLSIGFTLVIVFFVLVAILGSYVIYRPRGKQEKIGSLSCGVPFGFKIHGPGVP 4637
```

```
QY 137 ----- 136
Db 4638 TQSQVDHVLVRLNHPSEAPSPVAGDHMRPVGSHHLVGPPELLTKKFKPTAEMPQPOQ 4697
QY 137 ----- 136
Db 4698 QOQRQRPDIITERESPLIREDHHLPIPLHPLPLEHASSVDMGSEYPEHYDLENASSIAP 4757
QY 137 ----- 136
Db 4758 SDIDIVYHYKYGREAAGURKYKASVPPVSAYTHHKQNSGSOQOQOQOQHRHTAPFVTRNOG 4817
QY 137 ----- 136
Db 4818 COPPPPTTSRTHOSTPLARLSPSELSSQOPRILTLHDISGKPLQALLATTSSSGGV 4877
QY 137 ----- 136
Db 4878 GKDVHNSERSLNSPMSQLSQSSASROKPGVPOQAQOOTSMLGTABEIERLANGRPRT 4937
QY 137 ----- 143
Db 4938 CSLISTDAVSSSEAPRVSSSALHMSLGGVDVAHSTSTDSGNDSTFCSEIEYDNNSL 4997
QY 144 ----- 143
Db 4998 SCDGKYSTSKSLDGRSPVSRALSGGETSRNPPTTVTKTPIPPHAYDGFESSFRGSLST 5057
QY 144 ----- LGV 146
Db 5058 LVASDDDIANHLGSIYRKANGAAPSATTLGWYLLNMGPSYENLMGV 5105
```

RESULT 14
PKD1_HUMAN
ID PKD1_HUMAN STANDARD; PRT: 4303 AA.
AC P98161;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Polycystin precursor (Autosomal dominant polycystic kidney disease protein 1).
DE PKD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95254638; PubMed=7736581;
RA Gluecksmann-Kuis M.A., Tayber O., Woolf E.A., Bougueleret L.,
RA Deng N., Alperin G.D., Iris F., Hawkins F., Munro C., Lakey N.,
RA Duyk G., Schneider M.C., Geng L., Zhang F., Zhao Z., Torosian S.,
RA Reiders S.T., Bork P., Pohlschmidt M., Loehning C., Kraus B.,
RA Nowicka U., Leung A.L.S., Frieschaut A.-M.;
RT "Polycystic kidney disease: the complete structure of the PKD1 gene
and its protein.";
RL Cell 81:289-298(1995).
RN [2]
RP SEQUENCE OF 2769-4303 FROM N.A.
RX MEDLINE=94273192; PubMed=8004675;
RA Ward C.J., Peral B., Hughes J., Thomas S., Gamble V.,
RA Maccarthy A.B., Sloane-Stanley J., Buckle V.J., Kearney L.,
RA Higgs D.R., Ratcliffe P.J., Harris P.C., Roelfsema J.H.,
RA Spruit L.L., Saris J.J., Dauwerse H.G., Peters D.J.M.,
RA Breuning M.H., Nollst M., Brook-Carter P.T., Maheshwar M.M.,
RA Cordeiro I., Santos H., Cabral P., Sampson J.R., Janssen B.,
RA Hesselberg-Janssen A.L.W., van den Ouweland A.M.W., Eussen B.,
RA Verhoeve S., Lindhout D., Halley D.J.J.;
RT "The polycystic kidney disease 1 gene encodes a 14 kb transcript and
lies within a duplicated region on chromosome 16. The European
RT Polycystic Kidney Disease Consortium.";
RL Cell 77:881-894(1994).

[3] RT STRUCTURE BY NMR OF 275-354.
 RX MEDLINE=99107746; PubMed=9889186;
 RA Bycroft M., Bateman A., Clarke J., Hamill S.J., Sandford R.,
 RA Thomas R.L., Chochia C.;
 RT "The structure of a PKD domain from polycystin-1: implications for
 RT polycystic kidney disease.";
 RL EMBO J. 18:297-305(1999).
 [4] RT VARIANT ADPKD 3748-ARG--VAL-3752 DEL, AND VARIANT ASP-3632.
 RX MEDLINE=96108969; PubMed=8554072;
 RA Peral B., San Millan J.L., Ong A.C.M., Gamble V., Ward C.J.,
 RA Strong C., Harris P.C.;
 RT "Screening the 3' region of the polycystic kidney disease 1 (PKD1)
 RT gene reveals six novel mutations.";
 RL Am. J. Hum. Genet. 58:86-96(1996).
 [5] RT "Loss of heterozygosity in polycystic kidney disease with a missense
 RT mutation in the replicated region of PKD1.";
 RL Hum. Genet. 103:709-717(1998).
 [11] RT VARIANT ADPKD L-324 AND S-845, AND VARIANTS R-1399 AND L-1786.
 RX MEDLINE=99294580; PubMed=10364515;
 RA Thomas R.L., McConnell R., Whittaker J., Kirkpatrick P., Bradley J.,
 RA Sandford R.;
 RT "Identification of mutations in the repeated part of the autosomal
 RT dominant polycystic kidney disease type 1 gene, PKD1, by long-range
 PCR.";
 RL Am. J. Hum. Genet. 65:39-49(1999).
 [12] RT VARIANTS ADPKD P-2392 AND F-2423, AND VARIANTS R-1399; Q-2548 AND
 RP R-2638.
 RX MEDLINE=20046890; PubMed=1057909;
 RA Watnick T., Phakdeekitcharoen B., Johnson A., Gandalph M., Wang M.,
 RA Briefel G., Klinger K.W., Kimberling W., Gabow P., Germino G.G.;
 RT "Mutation detection of PKD1 identifies a novel mutation common to
 RT three families with aneurysms and/or very-early-onset disease.";
 RL Am. J. Hum. Genet. 65:1561-1571(1999).
 [13] RT VARIANTS ADPKD 3994-L--F-3996 DUPL; G-4136 AND C-4154, AND VARIANTS.
 RX MEDLINE=20441957; PubMed=10987850;
 RA Perrichot R.A., Mercier B., Simon P.M., Whebe B., Clodes J., Ferec C.;
 RT "DGE screening of PKD1 gene reveals novel mutations in a large cohort
 RT of 146 unrelated patients.";
 RL Hum. Genet. 105:231-239(1999).
 [14] RT VARIANTS ADPKD 3748-R--V-3752 DEL AND L-4132 DEL, AND VARIANT V-4045.
 RX MEDLINE=20112427; PubMed=10647901;
 RA Afzal A.R., Hand M., Ternes-Pereira E., Saggar-Malik A., Taylor R.,
 RA Jeffery S.;
 RT "Novel mutations in the 3' region of the polycystic kidney disease 1
 RT (PKD1) gene.";
 RL Hum. Genet. 105:648-653(1999).
 [15] RT VARIANTS ADPKD PRO-4225 AND TRP-4276.
 RX MEDLINE=99217041; PubMed=10200984;
 RA Badenas C., Torra R., San Millan J.L., Lucero L., Mila M.,
 RA Estivill X., Darnell A.;
 RT "Mutational analysis within the 3' region of the PKD1 gene.";
 RL Kidney Int. 55:1225-1233(1999).
 [16] RT VARIANTS ADPKD MET-2250 AND TRP-2329, AND VARIANTS CYS-2379; LEU-3066;
 RP VAL-3139 AND LEU-3193.
 RX MEDLINE=20311156; PubMed=10854095;
 RA Perrichot R., Mercier B., Quere I., Carre A., Simon P., Whebe B.,
 RA Clodes J., Ferec C.;
 RT "Novel mutations in the duplicated region of PKD1 gene.";
 RL Eur. J. Hum. Genet. 8:353-359(2000).
 [17] RT VARIANTS ADPKD PRO-2921 AND MET-3375, AND VARIANT LEU-3066.
 RX MEDLINE=20382887; PubMed=10923040;
 RA Koptides M., Mean R., Demetriou K., Constantiniides R., Pierides A.,
 RA Harris P.C., Deltas C.C.;
 RT "Screening of the PKD1 duplicated region reveals multiple single
 RT nucleotide polymorphisms and a de novo mutation in Hellenic
 RT polycystic kidney disease families.";
 RL Hum. Mutat. 16:176-176(2000).
 [18] RT VARIANTS ADPKD GLN-3719 AND PRO-3852, AND VARIANT VAL-4045.
 RX MEDLINE=20514565; PubMed=11058904;
 RA Agutari G., Savelli S., Garbo M., Bozza A., Augello G., Penolazzi L.,
 RA De Paoli Vitali E., La Torre C., Cappelli G., Piva R., del Senno L.;
 RT "Novel splicing and missense mutations in autosomal dominant
 RT polycystic kidney disease 1 (PKD1) gene: expression of mutated
 RT genes.";
 RL Hum. Mutat. 16:444-445(2000).
 [19] RT VARIANTS ADPKD SER-1166; GLU-1956; CYS-2408 AND GLY-2442--2443 INS,
 RP AND VARIANTS HIS-1995 AND ASN-2604.
 RX MEDLINE=20467506; PubMed=11012875;
 RA Phakdeekitcharoen B., Watnick T.J., Ahn C., Whang D.-Y., Burkhardt B.,
 RA Germino G.G.;
 RT "Thirteen novel mutations of the replicated region of PKD1 in an Asian
 RT population.";
 RL Kidney Int. 58:1400-1412(2000).
 [20] RT VARIANTS ADPKD TRP-3753 AND ASN-3815.
 RX MEDLINE=20275386; PubMed=10729710;
 RA Kim U.K., Jin D.K., Ahn C., Shin J.H., Lee K.B., Kim S.H., Chae J.J.,
 RA Hwang D.Y., Lee J.G., Namkoong Y., Lee C.C.;

RP "Novel mutations of the PKD1 gene in Korean patients with autosomal
RT dominant polycystic kidney disease";
RL Mutat. Res. 432:39-45(2000).
RN [21]
RP VARIANTS ADPKD Q-13; F-75; C-139; 1992-F-T-1993 DELINS L; 2220-R--P-
RP 2224 DEL; D-2336; D-2752; 2762-ILMR-2765 DUPL; M-2768; K-2771; P-2816;
RP S-2858; 3012-T-Y-3017 DEL AND 3748-L-R-3752 DEL, AND VARIANTS S-
RP 2674; M-2708; T-2734; L-2735; C-2765; M-2782; R-2814; G-2888; I-2905;
RP D-2966 AND L-3066.
RX MEDLINE=21063179; PubMed=1115377;
RA Rossetti S., Srzemecki L., Gamble V., Burton S., Sneddon V., Peral B.,
RA Roy S., Bakkaloglu A., Komel R., Winealis C.G., Harris P.C.;
RT "Mutation analysis of the entire PKD1 gene: Genetic and diagnostic
RT implications";
RL Am. J. Hum. Genet. 68:46-63(2001).
RN [22]

Query Match 26.9%; Score 208; DB 1; Length 4303;
Best Local Similarity 2.2%; Pred. No. 50;
Matches 87; Conservative 19; Mismatches 41; Indels 3862; Gaps 27;

QY 2 AGG-----ILH----- 7
DB 23 AGGPRCCGCEPPCLCGPAPGAACRVNCSGRGLRTGLPALRIPADATELDVSHLLRAL 82
QY 8 ----- 7
DB 83 DVGLLANLSALAELDISNNKISTLEEGIFANLNFNLSEINLSGNPFCCDCGLAWLPQWAE 142
QY 8 ----- 7
DB 143 QQVVVQPEAATCAGPGSLAGOPLLGTPLLDSGGBEYVACLPNSSGTVAAVFSAAHE 202
QY 8 ----- 7
DB 203 GLLOPEACSFCTSGGLAALSEQGWCLCGAAQPSASPACLSLCSGPPAPPAPTRCRGP 262
QY 8 ----- 7
DB 263 TLLQHVFPASGATLVCPHGLASGQLAAFHIAAPLVTDTRWDFDGSAAEVDAAAGPAAS 322
QY 8 -----LELL----- 11
DB 323 HRYVLPGRYHTAVLALGAGSALLGTDVQVEAAPALELVCPSVQSDSDLSIQNRGG 382
QY 12 ----- 11
DB 383 SGLEAAYSIVALGEEPAPRAVHPLCPDTEIFPGNGHCYRLVVEKAALQAEQOCQAWAGA 442
QY 12 -----VAVGP----- 16
DB 443 ALAMVDSPAVQRFVLSRVTRSLDWIGFSTVQGVGEPAPQGEAPFSLSCQNLPGEPHP 502
QY 17 ----- 16
DB 503 ATAHCVRLLGPTGWCNTDLCSPHSYVCELPQGGPVQDAENLLVGAPSGDLQGLTPLAQ 562
QY 17 -DVFOAQHEDTERVV-----LTNLNIGAEHLRDP----- 44
DB 563 QDGLSAPHEPVEVMVFPGLRLSREAFLLTAEFGTQELRRPAQLRQYRLLSTAGTPENG 622
QY 45 ----- 44
DB 623 SEPESSPDNRQTALPACMPGGRWCPGANICLPDASCHPQACANGCTSGPGLPGAPYAL 682
QY 45 ----- 44
DB 683 WREFLFSVPAGPPAQYSVTLHGQDVLMLPGDLVQLQHDAGPGALLHCSAPGHPGPRAPY 742
QY 45 ----- 44
DB 743 LSANASSWLPPLPAQLEGTWGCPACALRLLAQREQTLVLLGLRPNPGLRLPGRYEVRAEV 802

QY 45 ----- 44
DB 803 GNGVSRHNLSCSFDVVPVAGLRVIYPAPRDCRLVYPTNGSALVLQVDSGANATATARP 862
QY 45 --SLGAQPR-----VHLVKV----- 58
DB 863 GGSLSARFENVCPALVATFVPACPMETNDTLFSVVALPMLSEGEHVVDVVVENSASRANL 922
QY 59 ----- 58
DB 923 SURVTAEEPICGLRATPSPPEARVLOGVLVRYSPVVEAGSDMVFRMTINDKQSLTFQNVVF 982
QY 59 ----- 58
DB 983 NVIIQSAAVFKLSLTASNHVSNVTYNTVERMNRMOGLQVSTVPAVLSPNATLALTAG 1042
QY 59 ----- 58
DB 1043 VLVDSAVEAFLWTFDGEQALHQPFPYPNESFPVDPSPAQVLVEHNVTHTYAAPGEYL 1102
QY 59 ----- 58
DB 1103 LTVLASNAFENLTQQVPVSVRASLPNAVGVSDGVLVAGRPVTFYPHPLSPGGVLYTWD 1162
QY 59 ----- 58
DB 1163 FGDGSPVLTSQSPAANTHTYASRGTYHVRLEVNNTVSGAAQADVRVFEELRGLSVDMSLA 1222
QY 59 -----ILTEPE----- 64
DB 1223 VEQAGPVVVAAVOTGDNIWTTFDMGDGTVLSGPEATVEHVYLRAGNCTVTYVAGSPAGH 1282
QY 65 -----GA 66
DB 1283 LARSHLVLVLELVRVEPAACIPQPDARLTAYTGNPAHYLFDWTFDGSNTTVRGC 1342
QY 67 PNITANLTSS-----LLSVC----- 81
DB 1343 PTVTHNFRSGTFFPLALVLSRRAHVFTSICVEPEVGNVTLQPEROFVLGDEAWLVA 1402
QY 82 ----- 81
DB 1403 CAMPPFPYRYTWDFTBEAAPTRARGPEVTFIYRDPGSYLVTVTASNINISAANDSALVEV 1462
QY 82 ----- 81
DB 1463 QEPVLVTSIKVNGSLGLELOQPYLPSAVGRGRPASYLWDLDGCGWLEGPEVTHAYNSTGD 1522
QY 82 -----GW-----SQT----- 87
DB 1523 FTVRVAGNVEVSRSEALNVTVKRRVRGLVWNASRTVYPLNGSVSFSTSLGAGSDVRYSW 1582
QY 88 ----- 87
DB 1583 VLCDRCTPIPGPTISYTFRSVGTENIIVTAENEVGSQDSIFVYVYVQLIEGLVVGGR 1642
QY 88 ----- 87
DB 1643 YPPTNHTVQLQAVRDGTVNSYSWTAWDRDRGPALAGSKGSPSLTVLEAGTYHVQLRATNM 1702
QY 88 ----- 87
DB 1703 LGSAWADCTMDFEVPVGLMWAAVSPAAVNTSVTLAELAGSGGVYVWTSLEEGLSWET 1762
QY 88 ----- 87
DB 1763 SEPFTTHSFPTGLHLVTMTAGNPLGSANATVEVDVQVPVSGLSIRASEPGGSFVAAGSS 1822
QY 88 ----- 87
DB 1823 VPFWQLATGTVNSHCWAVPGSSKRGPHVTMVPDPAGTFSIRLNASNNAVSHVSATYNTLT 1882
QY 88 -----NPE----- 90

Qy	7	-----HLEL-----	10	Qy	35	-----	34
Db	416	NTGLISILEPVKQQAHFLEVTTSDRKASTKVLVKLGANSNPPEFTQAYKAAFDEN	475	Db	1496	SRDPLSLKFLRDPATGSLYTSKLOHEAVSPAHLTMVRDQDVVKRNFARIVVNVSDT	1555
Qy	11	-----LVAGPD-----	17	Qy	35	-----NIGABELLR	42
Db	476	VPIGTTIMSLSAVDPEDEGENGYVTSYIANLHVPPAIDHFTGAVSTSENLDYELMPRYT	535	Db	1556	NDIAPFTASSYKGRVYESAAVGSVVLQVLTALDKDKKNAEVLYSIESGHNIGNSFMI	1615
Qy	18	-----	17	Qy	43	DPSLGA-----	48
Db	536	LRIRASDWGLPYRREVEVLATITLNNLNDNTPLFEKINCEGTIPRDLCVGGEQITTVSAID	595	Db	1616	DPVLGSIKTAKELDORSQAQAEYDLWKATDKGSPMSEITSVRIFVTIADNASPKFTSKEY	1675
Qy	18	-----	17	Qy	49	-----	48
Db	596	ADELQVQYQIEAGNELDLFSLNPSNGVLSLKRSIMDGLGAKVSPHSLRITATDGENPAT	655	Db	1676	SVELSETVSIGSFVGMVMTAHSQSSVVEIKDGTGDAFDINPHSGTIIITQKALDFETLPI	1735
Qy	18	-----	17	Qy	49	-----QPRVHL-----	61
Db	656	PLYINITVAASHKLVNLQCEETGVAKMLAEKLLQANKLHQGEVEDIFFDSHVSNAHIPQ	715	Db	1736	YTLIIOGTNMAGLSTNTTVLVHLQDENNDNAPVFMQABEYTGGLISESASINSVVLTDNRNVL	1795
Qy	18	-----	17	Qy	62	-----EP-----	63
Db	716	FRSLPTGIVKENQPVGSSVIFMNSTDLDTGFNGKLVYAVSGNEDSCFMIDMETGMLK	775	Db	1796	VIRAADADKDSNALLVYHIVEPSTHTYFAIDSSGTGAIHTVLSLDYEETSIFHFTVOVHDM	1855
Qy	18	-----	17	Qy	64	-----EGAPNIT-----	70
Db	776	ILSPLDRETTDKYTLNITVYDLGIPQAAWRLLHVVVVVDANDNPPEFLOESYFVEVSEDK	835	Db	1856	GTPLRAEYAAANTVHVHIDINDCPPVFAKPLYEASLLLPYKGVKVITVNATDADSAFS	1915
Qy	18	-----	17	Qy	71	-----ANLTS-----	75
Db	836	EVHSEIIQVEATDKDLGNHGVTSYILTDITFSIDSVTGVVNIARPLDRELQHEHSLKI	895	Db	1916	QLIYSITEGNEGKFSMDYKGTGALTQVNTTQLRSRYELTVRASDGRFAGLTSVKINVKES	1975
Qy	18	-----	17	Qy	76	-----	75
Db	896	EARDOAREEPOLFSTVVVVKVSLSDVNDNPPTFIPNVYRKVREDLPNECTVIMMLEAHDPD	955	Db	1976	KESHLKFTQDVYSAVVKENSTEATLAVITAIGSPINEPLFYHILNPDPRFKISRTSGVL	2035
Qy	18	-----	17	Qy	76	-----	75
Db	956	LGQSGOVRYSLLDHGEGNFVDVLDKSGAVRIVQQLDFEKKQVYNLTVRAKDKKPVLSST	1015	Db	2036	STGTTPDREQEAFDWWVEVIEBEHKPSAVAHVVVKVIVEDQNDNAPVFNLPYAVVKV	2095
Qy	18	-----VF-----QAHOED-----	25	Qy	76	-----	75
Db	1016	CYVEVEVDNENLHPVPFSSFVEKGTVKEDAPVGSVMTVSAHDEADAGRDEIRYSIRD	1075	Db	2096	DTEVGHVIRYVTAVDGRNGEVHYLKEHHEHFOIGLGEISLKKQFELDTLNKEYLV	2155
Qy	26	-----TERYVLT-----	32	Qy	76	-----	75
Db	1076	GSVGVFKEETGVIETSDRLDRESTSHYLTVFATDQGVVPLSSFIEIYIEVEDVNDN	1135	Db	2156	TVVAKDGGNPAFAEVIPIVITVMNKAMPVEKPFYSAETASIOVHSPPVHVQANSPEGL	2215
Qy	33	-----	32	Qy	76	-----	75
Db	1136	APQTSPEVYYPEIMENSPKDVSVVQIEAFDPDSSNDKLMYKITSGNPQGFPSHPKTGL	1195	Db	2216	KVFYSITDGDPPFSQFTINFNTGVINVIAPLDPEAHPAYKLSIRATDSLTCAHAEVFVDII	2275
Qy	33	-----	32	Qy	76	-----	75
Db	1196	ITTSRKLDRQODEHILEVTVDNGSPPKSTIARVIVKILDENDNKPQFLQKFKIRLP	1255	Db	2276	VDDINDNPPVFAOQSYAVTLSEASVIGTSVVQVRATDSDBEPNRGISYOMFGNHSKSHDH	2335
Qy	33	-----	32	Qy	76	-----	75
Db	1256	EREKPDRENNARREPLYRIATDKDEGNABEISYIEDGNEHGKFFIEPKTGTVVSKRFS	1315	Db	2336	FHVDSTGLISLRLTLDYEOSROHTTFVRAVDCGMPTLSSDVIIVDVTDLNGNPPLPEQ	2395
Qy	33	-----	32	Qy	76	-----	75
Db	1316	AAGEYDILSIKAVONGRPQKSTTRLHIEWISKPKQSLPISFEESFTFTVMESDPVAH	1375	Db	2396	QIYEARISEHAPHGHFVTCVKAYDADSSDIDKLOYSILSGNDHKHFVIDSATGIITLSNL	2455
Qy	33	-----NL-----	34	Qy	76	-----	75
Db	1376	MIGVISVEPPGIPLWFEDITGNGYDSHFDVDKGTGTIIIVAKPLDAEQKSNYLTVEATDGT	1435	Db	2456	HRHALKPPYSLNLSVSDGVFRSSSTQVHVTVIGGNLHSPAFLONEYEVELAENAPLHTVM	2515
Qy	35	-----	34	Qy	76	-----	75
Db	1436	TTILTQVFIKVIDNDRHROPFSTSKYEVVIPEDTAPETELIQISAVDQDEKGNKLIYTLOS	1495	Db	2516	EVKTTDGDGSIYGHVTHYIHNDFAKDRFYINERGOIFTEKLDRETPAEKVISVRLMAKD	2575
				Qy	76	-----	75

Db	2576	AGGKVAFTVNVILTDNDNAPQFRATKYEVNIGSSAAKGTSWKASDADEGSNADITY	2635	Db	3656	EFVGDYWRNFORALRNILGVRNRNDIQIVSLOSSEPHPLDVLVLFVEKPGSAQISTKOLLH	3715
Qy	76	-----	75	Qy	106	-----	105
Db	2636	AIEADSESVKENLKNLGSVITTKESLIGLENEFFTFVRAVDNGSPSKESVVLVYVKI	2695	Db	3716	KINSSVTDIEEIIIGVRILNVFQKLCAGLDCPWKFCDEKVSVDSEVMSTHSTARLSFVTPR	3775
Qy	76	-----	75	Qy	106	-----	105
Db	2696	LPPEMQLPKFSEBFFVFTVSEDPVPGTEIDLIRAEHSGTVLSLVKGNTPESNRDESFI	2755	Db	3776	HRAAVCLCKEGRCPVHHGCCDDPCPEGSECVSDPWEKHCTVCPCPSGRFOCGSSSMT	3835
Qy	76	-----	81	Qy	106	-----	109
Db	2756	DRSGRLKLEKSLDHEHTKWQFSILARCTQDDHEMVASVDVSIQVKDANDNSPVFESSP	2815	Db	3836	LTGNSYVKYRLTENENKLEMKLTWRLRTYTHAVVMYARGTDYSILEIHHGRLQVKKFCG	3895
Qy	82	-----	83	Qy	110	-----	109
Db	2816	YEAFIGENLPGSRVQIRASDADSGTNGQVMYSLDOSQOSVEVIESFAINMETGMITLIK	2875	Db	3896	SGPGIVSQSTQVNDGQWHA VALEVNGNYARLVLDQVHTASGTAPGTLKTLNLDNYVFFG	3955
Qy	84	-----	83	Qy	110	-----	116
Db	2876	ELDEKRDNYQIKVVASDHGEKIQLSSTAIVDVTVDVNDSPRFTAEIYKGTVSEDDPO	2935	Db	3956	GHIRQOQTRHGRSPQVNGFRGCMDSIYLNGOELPLNSKPRSYAHIEESVDVSPCCFLTA	4015
Qy	84	-----	88	Qy	117	-----	116
Db	2936	GGVIALSTTDADSEINRQVTFITGDPGLQFAVETIQNEWKVYVKKPLDREKRDNYL	2995	Db	4016	TEDCASNPONGGVCNPNPAGYYCKSALYIGTHCEISVNPCSSNPCLYGTCVVDNGG	4075
Qy	89	-----	91	Qy	117	-----	128
Db	2996	LTITADGTFFSKAIVEVKVDANDNSPVCKEKLTSYDITPDEVLPGLKIMQISATDADIR	3055	Db	4076	FVOCRGLYTQRCQLSPYCKDEPKNGGTCFDSLDGAVCOCDSGFRGRCQSDIDECSG	4135
Qy	92	-----	91	Qy	129	-----	128
Db	3056	SNAEITYTLGSGAEKFLNPDTGELKSTPLDREBQAVYHLLVRATDGGRCQASIVV	3115	Db	4136	NPCLHGALCENTHSGYHCNCSHEYRGHRHCEADAPQYVSTPMNIGLAEGIGIVFVAGIF	4195
Qy	92	-----	96	Qy	129	-----	128
Db	3116	TLEDVNDNAPEADPYAITVENTEPGTLTRVQATDADAGLNKILYSLIDSADGQS	3175	Db	4196	LLVVVFLCRKWI SRKKKHQAEPKDHLGLPATAFLORPYFDSKLNKNIYSIDIPQVVRP	4255
Qy	97	-----	96	Qy	129	-----	128
Db	3176	INELSGIIQLEKPLDRELOAVTSLKAVDOGLPRRLTATGTIVSVLNDINDNPPVFEYR	3235	Db	4256	ISYTPSIPSDSRNLDNRNSFEGSAIPEHPEFTFNPESVHGHRKAVCVSVAPNLP PPP	4315
Qy	97	-----	96	Qy	129	-----	132
Db	3236	EYGATVSEDILVCTEVLQVVAASRDIEANAEITYSIISGNEHCKFSDSKTGAFFIENL	3295	Db	4316	SNSPSDSDSIQKPSHDFDYDTKWVDLPCLSKKPLEEKSPQYSARESLSVQSLSFQS	4375
Qy	97	-----	96	Qy	133	-----	144
Db	3296	DYESSHEYVLTVEATDGGTSLSDVATVNVNVTDINDNTPVFSQDTYTTVISEDVLEQS	3355	Db	4376	ESCDNGYHWDTSOMMPSVPLDIOEPFNVEVIDEOTPLYSADPNADTDYYPGGYDIES	4435
Qy	97	-----	96	Qy	145	-----	144
Db	3356	VITVMADADAGPSNHIHYSIIDNCGSFTIDPVRGEVKVTKLLDRETISGYTLVQAS	3415	Db	4436	DEPPPPEDFPADELPLLPPEFSNQFESIHPPRDMPAAGSLGSSSRNRQRPNLQVLPNF	4495
Qy	97	-----	96	Qy	145	-----	144
Db	3416	DNGSPRVNTTVNIDVSDVNDNAPVFSRGNYSVIIQENKPVGFSVLQVLTDEDSSHNG	3475	Db	4496	YPLDMSEPTKTGTGENSTCREPHAPYPPQYQHRHEAPAVESMPMSVYASTASCDSVSACC	4555
Qy	97	-----	96	Qy	145	-----	148
Db	3476	PPFFFTVTCNDEKAFENVPGVLLTSSAIKRKEKDHYLLQVKVADNGKPOLSSLTYIDI	3535	Db	4556	EVESEVMMSDYESGDDGHFEVTI	4579
Qy	97	-----	96				
Db	3536	RVIEESIYPPAILPLEIFITSSGEYSGVIGIKIHATDQDVDTLTYSLDPQMDNLFVS	3595				
Qy	97	-----	96				
Db	3596	STGGKLIHKKLDIGVLLNVSVTDGKFTTVADITVHIROVTOEMLNHTIAIRFANLTPE	3655				
Qy	97	-----	105				

Search completed: March 20, 2003, 12:39:21
Job time : 75 secs

Query Match	34.2%	Score 264;	DB 2;	Length 15281;
Best Local Similarity	0.8%	Pred. No. 0.013;		

Matches 102; Conservative 24; Mismatches 21; Indels 12475; Gaps 32;

Qy	1	AAGGIL	19	-----	18
		:			
Db	2098	AAGHLEIGTGTGMLNQLGAGLSYIGLEPSQSAQVFNKAAQTFFGLEGAQHVGHT	3178	PIEGMAEETIRITEILADAKTDDINGLAASQPTAASLAYVFTSGSTGRPKGVMEHR	3237
Qy	7	-----	19	-----	18
Db	2158	AMDTGRLSALPDIVINSVAQYFSPSREYLAEVVEALVRIPGVRRIFFGDMRTYATHKDF	3238	GIVRLTKQNTSKLPESFMAHISNLAFDASVWEVFTLLNGGTLVCIDYFTLLESTAL	3297
Qy	7	-----	19	-----	18
Db	2218	LVARVHTNGSKVTRSKVOEVARLEEEELVDPAPFTSLKESLSEIEHVEILPKNM	3298	EKVFFQORVNVALLPALLKQCLDNSPALVKLTSLVLYIGGDRLDASDAKARGLVQTAQF	3357
Qy	12	-----	19	-----	18
Db	2278	KVNNELSYRYGAVLHNRHNQNSRSIHKINAESWIDFASSQMDROGLARLLKENKDAE	3358	NAYGPTENTVMSTIYPIAEDPPFINGVPIGHAVSNSGAFVMDQNOQITPPGANGELIVTGD	3417
Qy	12	-----	19	-----	18
Db	2338	STAVFNI PYSKTIVERHTAKSLADHDGDDTHSSIDGVAMISAAREKASQCPSLDVHDLV	3418	GLARGYTTSSLNTGRFINVDIDGEOVRAYRTGDRVRYRPKDLQIEFFGRIDHOVKIRGHR	3477
Qy	12	-----	19	-----	18
Db	2398	QLAEDAGPRVEVSWARQSONGALDVFFHHFQPTENESRALVDPTDYKGOQARS LNRP	3478	IEPAEVEYALLSHDLVTDAAVVTHSQENQDLEMVGFAARVADVREDESSNOVOEQTHF	3537
Qy	12	-----	19	-----	18
Db	2458	LORVESRRIEAQVRELOVLLPAYMIPARIIVLQNMPLNTSGKVDKELTLRAKVAART	3538	DSIAYADITTTDOQSLGRDFMSWTMYDGS LIKKSQMOEWLDDTMRSLLDSQPPGHVLEV	3597
Qy	12	-----	19	-----	18
Db	2518	PSELVAPRDSIEALICEFKDVLGVEGVTDNFNFGVGHSLATKLAARLSQLNAQIA	3598	GTGTGMLNLCREGGLOSYYGLEBPSPSATAFVNKAASFPGLIEDRIRVEVGTATDRL	3657
Qy	12	-----	19	-----	18
Db	2578	VKIDFRPVIADLAATIQDTEHNPIPTSYTGPVEQSPAQGRMLFDQLNVGATWYLM	3658	GDDLHAGLVVNSVAQYFSPQDYLAQLVRDLTKVPGVERIFFGDMRSHAINRDFVARAV	3717
Qy	12	-----	19	-----	18
Db	2638	PFVRLRGPLVVSALAAALLALEERHETLRTTFIEQEGIGMQVIHPFAPKELRVIDVSGE	3718	HALGDKATKAEIQREVVNRMBEDELVDPAFTSLTTOVENIKHVEILPKRMATNELS	3777
Qy	12	-----	19	-----	18
Db	2698	EESTIQKILEQOTTPFNLAEPGFRLLALKGTGEDEHILSTVMHHAISDGWSVDIPQEI	3778	SYRYAHLVNDLAKPAHKVSPGAWVDAATKMDRDALIRLLRGTKISDHAIATIPNSK	3837
Qy	12	-----	19	-----	18
Db	2758	QFYSAILRGHDPLAQIAPLSIQYRDFATWQRQIQVAEHRRQLAYWTQKADNKP AELL	3838	TIVERTICESVYDLGGDAKOSNDRSVLSAARSNAVKVASLSAIDLVDIAQEAQFRVEIS	3897
Qy	12	-----	19	-----	18
Db	2818	TDFKRPMLSGRAGEIPVVVDGLIYEKLQDFCRIRQVTAFTVLLAAFRAAHYRMTGTEDA	3898	CARQWSQNGALDAVFHHLGPSPOSSHVLIDFLTDHOGRPBEALTNHPLHRAQSRRVERQI	3957
Qy	12	-----	19	-----	18
Db	2878	TICTPIANRNRPELEGLIGFVNQTCMRITVDVEDSFETLVHQVRETTLAAHANQDVPE	3958	RERLOTLLPAYMTPAQIMVLDKPLNANGKVDKOLTORAQTVPKAKQVSAPVAPRTEIE	4017
Qy	12	-----	19	-----	18
Db	2938	QIVSNILPGSSDTSRNPVLQMPALHSONQNLGKVRLEGIEEIIISIAETTRFDIEFHLYQ	4018	RVLCOEFSVDVLGVDIGIMENFPDLGCHSLMATKLAARISRRLETHSVSVKEIFDHPRVCDL	4077
Qy	12	-----	19	-----	18
Db	2998	EAERLNGSIYVAADLVFPETIQSVITIFOGILQKGLGEPDMPVASKALDGGLESURSTGL	4078	VLIVQGSAPHDPVSTKYTGPVQSFQAGRLWFLDQLNFGATWYLMPLAVRLRGMNVH	4137
Qy	12	-----	19	-----	18
Db	3058	LHPQOTDPCDASVVOIFKQQVAVNPDVIAVRDESTRLSYADLRKSDQVACWLSRRGIA	4138	ALTAALLALERHELLRTTFYEONGVGMQVNPVVTETLRIIDLNSGDGDLPTLKKEQT	4197
Qy	19	-----	19	-----	18
Db	3118	PETFAVILAPRSCETIVAILGLVKANLAYLPDVNVNPAASLEAILSEVSGMLVLVGAET	4198	APFHLETPGRVALLRLPGDYILSVVMHHIISDGWSVDVLFOELGQFYSTAVKGDHPL	4257

```
QY 19 ----- 18
Db 4258 SQTPLPIHYRDFALWQKPTQSEHERQLQYVWEQLVDSAPAELLTDLPRPSILSQAG 4317
QY 19 ----- 18
Db 4318 EMSVTIEGALYKNLEEFRCVRHVRTSFVLLAALRAAAHYRLTGSEDATIGTPIANRRPEL 4377
QY 19 ----- 18
Db 4378 EQIIGFVNTQCIRITVNEDETFESLVQVVRSTATAAFAHQDVPFKEIVSTLLPGSRDAS 4437
QY 19 ----- 18
Db 4438 RNPLVQLMFAVHSQKNLGELKLENAHSEVVPTBITTRFDLEPHLFQDDKLEGSILYSTD 4497
QY 19 ----- 18
Db 4498 LFEAVSQSLLSVFQILRRGLNGPDVPISTPLQDGIVDLQORQGLLDVQKTEYPRDSSV 4557
QY 19 ----- 18
Db 4558 VDFVHEQVSINPDSIALIHGSEKLSYAQLDRESDRVRLRHRSFSSDTLIAVLAPRSC 4617
QY 19 ----- 18
Db 4618 TIIAFLGILKANLAYPLDVKAPAAIDAIVSSLPGNKLILIGANVTTPKLOEAAIDFVP 4677
QY 19 ----- 18
Db 4678 IRDTFTLTLDGTLDQPTIERSAQSLAYAMFTSGTRPKGMVQHRNIVRLVKNVNV 4737
QY 19 -----FOAHQ----- 23
Db 4738 AKOPAAARIAHISNLAFDASSWEIYAPLLNGGAIVCADYFTIDPQALQETFOHEIRGA 4797
QY 24 ----- 23
Db 4798 MLPPSLLKQCLVQAPDMISRLDILFAAGDRFSSVDALQORLVGSGVFNAYGPTENTILS 4857
QY 24 ----- 23
Db 4858 TIYNVAENDSFVNGVPIGSAVNSGAYIMDKNQQLVPAGVNGELVVTDGLARGYMDPKL 4917
QY 24 ----- 23
Db 4918 DADRFIQLTVNGSEQVRAYRTGDRVRYRPKDFQIEFFGRMDQQIKIRGHRIPAEVQAF 4977
QY 24 ----- 23
Db 4978 LNDGFVEDVAIVIRTPENQPEMVAFVTAKGDNAREEEATTOIEGWEAHFEGGAYANIE 5037
QY 24 -----EDT----- 26
Db 5038 EIESEALGYDFMGWTSMDYDGTIEDKDDEMRLNDTMRSLLDGKPAGRVLEVTGTGIMF 5097
QY 27 -----ERV----- 30
Db 5098 NLGRSQGLERYIGLEPAPSAAEFVNNAKSPFGLAGRAEVHVGTAADVGTLOGLTSDMAV 5157
QY 31 ----- 30
Db 5158 INSVAQYFPTPEYLAETIKSILVQVPGMKRIYLGDMRSMWNRDFAAARAAYSADNASKD 5217
QY 31 ----- 30
Db 5218 RVROMMELEEBELVDPAFFTALASQLQDRHQHVEILPKRMKATNELSSRYAAVLH 5277
QY 31 ----- 30
Db 5278 ISDEPLIYKIDPEWINFEGSRLTREALAQVLKENENAESVAISNIPYSKTVERHIVR 5337
QY 31 ----- 30

Db 5338 SLDOEDANAPESMDGSDWISAVTRAQOCHTLSASDLFDAEDAGFRVEVSWARHSQH 5397
QY 31 -----LTN----- 33
Db 5398 GALDAVFHHLKATEDSRVLKFKPTDHQGRPLKSLTNQPLLPAQSRRAELLIREGLOTL 5457
QY 34 ----- 33
Db 5458 PPYMIPSOITLIDRMPLNANGKVDRELARRAKITOKSKPVEDIVPPRNSVEATVCKGFT 5517
QY 34 ----- 33
Db 5518 DVLGVEGVTDNFPNLGCHSLMATKLAARLGRQLNTRISVRDVPDQPVADLAAVIORNS 5577
QY 34 -----LNIGA----- 38
Db 5578 APHEPIKADYTGVPVPOSFAQGRQLWFLDQLNVGATWYMLPGIRLHGSLRVDALATAISA 5637
QY 39 ----- 38
Db 5638 LEORHEPLRTTFHEEDGVGVVQVQDHRPKDLRIIDLSTQPKDAYLAVLKHEOTTLFDLAT 5697
QY 39 ----- 38
Db 5698 EPGMRVALIRLGBEEHLSIVMHHIISDGSVEVLFDEMHRYFSSALROODPMEOILPLP 5757
QY 39 ----- 38
Db 5758 IOYRDAFAMQKTEQVAEHQRQDYWTEHLADSTPAELLTDLPRPSILSGRANELPLTIE 5817
QY 39 ----- 38
Db 5818 GRLLDKLRAFRCRVHQATPFVILLALRAAAHYRLTGAEDATLGTPTIANRRPELENMIGFF 5877
QY 39 ----- 38
Db 5878 VNTQCMRIAIEENDNPESLVRVRSTATSAFANQDVPPESIVSSLLPGSRDASRNPLVQV 5937
QY 39 ----- 38
Db 5938 ILAVHSQDLGKLTLEGLRDEAVD9A18TRFDVEFHLPEHADRLSGSVLYAKELFKLRTI 5997
QY 39 ----- 38
Db 5998 ESVWSVFLETLRRALDQPLTPLAVLPLTDGVGEIASKGLLDVPRDTPRDANIVEVFOOH 6057
QY 39 ----- 38
Db 6058 VRATPDIAVKDATSILTVAQLDQOSDRLAIWLSRRHMPETLVGLVLAAPRSCETIIAMFG 6117
QY 39 ----- 38
Db 6118 IMKANLAYPLDINSIPAARLRSILSAVDGNKVLVLLSGVGTAPQENPEVEAVGIOELIAG 6177
QY 39 ----- 38
Db 6178 TGLDKTQGSNARPSATSLAYVIFTSGTGKPKGMVVEHRSVTRLAKPSNVISKLPQGARV 6237
QY 39 ----- 38
Db 6238 AHLANIAFPDASIMEIATTLNGATLVCLDYHTVLDCTRLKEVPERESITVVVTLMPALLKQ 6297
QY 39 ----- 38
Db 6298 CVAEIPETLAHLDLLYTGDRVGGHDAMRARSVLKIGMFCSGYGTENTVISTIVEYDADE 6357
QY 39 ----- 38
Db 6358 MFVNGVPIGKTVSNGAYVMNRQQLVPSGVVGLVVTGDLARGYTDPSLNKURFIYIT 6417
QY 39 ----- 38
```

```
Db 6418 VGESIRAYRTGDRVRPHDLQIEFFGRMDQVKIRGHRIBEPGEVESALLSHNSVQDAA 6477
Qy 39 ----- 38
Db 6478 VVICAPADQSGAEMVAFVAARNTEDETOEEBAVQVQGWETHFETAAYSEVKIROSE 6537
Qy 39 ----- 38
Db 6538 VGNDFMGWTSMYDGEIDKTMHEWLNDRMILDAEPGHVLEIGTGTGMVFNLAACP 6597
Qy 39 ----- 38
Db 6598 GLOCYVGFEPSKAAQFVNDAAQSPALKDGRSIVHVGTAIDINKAGPIQPRLVINSVA 6657
Qy 39 ----- 38
Db 6658 QYFPTPEYLFRRVVEALVQIPSVVERIVFGDMRTNAINRDFVASRALHTLGEKANKRLVROM 6717
Qy 39 ----- 45
Db 6718 IVELEANEBELLTDPAFFTSRLRLGEKIKHVEILPKTKATNELSKRYAAVLHVGRSR 6777
Qy 46 ----- 45
Db 6778 EOSTIHQSPNAWIDFAADGLDROT LINLKEHKDAGTVAIGNIPYSKTIVERFVNKSL 6837
Qy 46 ----- 45
Db 6838 EDDMEGQNSLDGSAWAAVRMAQSCPSLDAMDVKEIAQAGYQVEVSNAROWSONGAL 6897
Qy 46 ----- 49
Db 6898 DAIFHFFPPKEGARTLIEFPTDYEGRNVTLTNRPLNSIQSRRLGTQIREKQLTLLPPY 6957
Qy 50 ----- 49
Db 6958 MIPSRIWLDQMPVNNCKIDRKELVRRAIVAPKPSAATRAVPRNIEAILRDEFEDVL 7017
Qy 50 ----- 49
Db 7018 GTEVSLDNFFDLGGHSLMATKLAARVSRRLDAHISIKVDQPVLDLAASIQRESAPH 7077
Qy 50 ----- 49
Db 7078 EPIQRPYTPAEQSPAGRLWFLDQNLGATWYLMPLAIRGQLRVAALSAAALFALER 7137
Qy 50 ----- 49
Db 7138 RHETLRTTFBESDGVQIVGEARNSDLRVHDVSTGDDGEYLEVLRREQTVPFDLSSEPG 7197
Qy 50 FRVHLVK----- 56
Db 7198 WRVCLVKTGEEDHVLIVMHHIIYDGNVSDILRGELGFYSAAALRGQDPLHANPLIQY 7257
Qy 57 ----- 56
Db 7258 RDFAAWQREAKOVEEHORQLGYWSKQIVDSTPAELLTDLPSPILSGRAGSVDDVTIEGVS 7317
Qy 57 ----- 61
Db 7318 YGALQSFCTRSTVTFWLLTVFRIAHFRLTAVDDATICTPIANRRNPELETLVCCFVNT 7377
Qy 62 ----- 61
Db 7378 QCMRISIADDDNFEGLRVQVRNVATAAYANQDVPFERIVSALVPGSRNTRSNPLVQLMFA 7437
Qy 62 ----- 61
Db 7438 VOSVEDYDQVRLEGLSVMPGEASTREDMEPHLVPGDKLTGSLVLYSSDLFEOGTQNF 7497
Qy 62 ----- 61
Db 7498 VDIQFQCLRSVLDOPLTPISVLPSFNAISNLESLLDLEMTSDYPRORTVVDFREQAAI 7557
Qy 62 ----- 61
Db 7558 CPDSIAVKDSSQLTYAQLDEQSDRVAAMLHERHMPAESLVGLSPRSCETIIAYFGINK 7617
Qy 62 ----- 61
Db 7618 ANLAYPLDVYAPDARLAAILOTVEGERLLLLGAGVPOPGIOIPLRLSTAYIAEALSHATT 7677
Qy 62 ----- 65
Db 7678 VDVTSIQPSATSALAVIIFTSGTGKPKGVMIEHRGIVRLVRDNTVNVPPESGSLPVSH 7737
Qy 66 ----- 65
Db 7738 FSNLAWDAATWEIYTAVLNGGTVCIDRDTMLDIAALNSTFRKENVRAAFPTAPLKQCL 7797
Qy 66 --APNITANL----- 73
Db 7798 AETPELVANLEILHTAGDRLOPDGDANLAGKTAKGGIFNVLGHTEHTAYSTFYPVVGEEFT 7857
Qy 74 ----- 73
Db 7858 VNGVPVGRGISNSHAYIIDRHQKLVPAQVMGELILTDGVARGVYDTSALNKRFEVYIDIN 7917
Qy 74 ----- 73
Db 7918 GKSTWSYRTGDKARYPRPDGQLEFFGRMDQVVKIRGVRIBEPGEVELTLLOHKSVLAAATV 7977
Qy 74 ----- 73
Db 7978 VRRPPNGDPEMIATITDAEDDVOTHKAIYKHLQGLPAYMIPSHLVILDOMPTDNGKV 8037
Qy 74 ----- 73
Db 8038 DRKDALARQTVQKRSSTAARVPRDEVEAVLCEEYSNLLEVEVGITDGFDLGCHSLA 8097
Qy 74 ----- 73
Db 8098 TKLAARLSROLNTRVSKVDQDQILADLADIIRRGSHRHDPIPATPYTGPVEQSPAQR 8157
Qy 74 ----- 73
Db 8158 LWFLEQLNLGASWYLPFAIRMRGPIQTKALAVALNALVHRHEALRTTFEDHGVGVQVI 8217
Qy 74 ----- 73
Db 8218 QPKSSQDLRIIDLSDAVDDTAYLAALKREOTTAFDLTSEPGWRVSVLLRLGDDDDYILSIVM 8277
Qy 74 ----- 73
Db 8278 HHIIISDGTVDVLRLOELGOFYSAAIRGOEPLSOAKSLPIQYRDFAVMQRQENQIKEQAK 8337
Qy 74 ----- 73
Db 8338 LKYSQQLADSTPCEFLTDLPSPILSGEADAVPMVIDGTVYQLLTDFCRTHQVTSFVL 8397
Qy 74 ----- 73
Db 8398 LAAFRTHVRLTGTILDATVGTPIANRRNPELEGLIGFFVNTQCMRMAISETETFESLVQ 8457
Qy 74 ----- 73
Db 8458 VRLTTTEPANQDVPFEQIVSTLLPGSRDTSRNPVLQVMFALQSOQDLGRIGLEGMTDEA 8517
Qy 74 ----- 73
Db 8518 LETPLSTRIDLEVLHFOEVBKLSGLSYSTDLFEVETIRGIVDVLETLRRGLEQPKORL 8577
Qy 74 ----- 73
Db 8578 NAMPITDGIKLRDOGLLTVAKPAYPRESSVIDLFROOVAAPDAIAWDSSTLTLYADL 8637
```

```
QY 74 ----- 73
Db 8638 DQCSNLAHLWCORNMAPETLVAVPAPRSLTIIVAFGLVKANLAYLPLDVNAPAARIEA 8697
QY 74 ----- 73
Db 8698 ILSAVPGHKLVLVOAHGPELGLTWADTELVOIDEALSSSGDHEQIHASGPTATSLAYV 8757
QY 74 ----- 73
Db 8758 MFTSGSTGPKGVMDHRSIIRLVKNSDVVATLPTPRMANVSNLAFDISVQIYTALLN 8817
QY 74 ----- 73
Db 8818 GGTIVCLDYLTLLDSKILYNVFEAQNAAFTBVLLKQCLGNPAIISRLSVLPNVGR 8877
QY 74 ----- 73
Db 8878 LDAHDAVASGLIODAVYNAYGPTENGQMSTMYKVDVNEFPVNGVPIGRSITNSGAYVMD 8937
QY 74 ----- 73
Db 8938 GNOOLVSPGVMEIIVTGDGLARGYTDSALDEDRFVHVTTIDGEENIKAYTGDVRYRPK 8997
QY 74 ----- 73
Db 8998 DFEIEPGRMDQOVKIRGRIEPAEVEHALLGHDVHDAAVLVRKPAQEPENIAFITSQ 9057
QY 74 ----- 73
Db 9058 EDETIEQESNKQVQMGHEFVSRYADIKDLDTSTFGHDFLGWTSYDGVDPVNEKME 9117
QY 74 ----- 78
Db 9118 WLDETTASLLDNRPBGHILEIGAGTGWILNGLKVDGLQYVGLDPAAPSAIIFVNEAVKS 9177
QY 79 ----- 78
Db 9178 LPSLACKARVLVGTALDIGSLDKNELQPELVVINSVAQYFPTSEYLIKVKVAVVEVPSVK 9237
QY 79 ----- 78
Db 9238 RVFFGDIRSOALNRDPLAARAVRALGDNASKQIREKIAELESEEEELLVDPAFVSURS 9297
QY 79 ----- 78
Db 9298 QLPNIKHVEVLPKLMKATNELSSRYAAVLHISHNEEBQLLIQIDPTAWVDFAATQKDS 9357
QY 79 -----SVC 81
Db 9358 OGLRNLQOGRDDVMIAVGNIPYSKTIIVERHIMNSLDQDHVNSLDGTSWISDARSAAAIC 9417
QY 82 ----- 81
Db 9418 TSFDAPALTQAKEGFRVELSWARQSONGALDAVFHRLATDANCERSVLVHFPTDQ 9477
QY 82 ----- 81
Db 9478 GRQLRTLNRPLQRAQSRRIESQVFEALQTPALPAYMTPSRIIVLPQMPTNANGKVDKQL 9537
QY 82 ----- 81
Db 9538 ARRAQVVAKRKAVSARVAPRNDTEIVLCEEYADILGTEVGTIDNFFDMGCHSLMATKLAA 9597
QY 82 ----- 81
Db 9598 RLSRRLTRVTVEFKPVLAJADLAASIEQGSTPHLPFIASSVSYSGPVEQSYAQRLWFLD 9657
QY 82 ----- 81
Db 9658 QFNLNATWYHMSLAMRLGLPLNMDALDVALRALEQRHETLRTTFFBAQKDIGVQVVEAGM 9717
QY 82 ----- 81

QY 9718 KRLKVLDSLKDNEKEHMAVLENEQMRPFTLASPCWKGLARLGPTPEYILSLVMHMFSD 9777
QY 82 GWS ----- 84
Db 9778 GNSVDILRQELGQFYSAALRGDRDPLSQVKPLPIQYRDFAAWQKEAAQVAHEROLAYMEN 9837
QY 85 ----- 84
Db 9838 QIADSTPGEILLTDFRPQFLSGKAGVIPVTIEGVPVYEKLLKFSKERQVTLFSLVLLTAFRA 9897
QY 85 ----- 84
Db 9898 THFRLTGAEDATICTPIANRNRPELEHIIIGFFVNTOCMRLLLDTGSTFESLVQHVRSVAT 9957
QY 85 ----- 84
Db 9958 DAYSNQDIPFERIVSALLPGSRDASRPLIQLMFALHSQDPLNITLEGLEHERLPTSA 10017
QY 85 -----OTIN----- 88
Db 10018 TRFDMEPHLFQEPNKLSGSILFADELFPQPETINSVVTVFOEILRRGLDQPOVSISTMPLT 10077
QY 89 -----PE----- 88
Db 10078 DGLIDLEKLGLLBIESSNPPRDYSVVVDVPRQOVAANPNAPAVVDSETSMTSLDQKSEQ 10137
QY 89 -----PE----- 90
Db 10138 IAAWLHAQGLRPESLICVMAPRSFETIVSLFGILKAGAYALPLDVNSPAARIQPILSVE 10197
QY 91 ----- 90
Db 10198 GKRLVLLSGIDMPOSMDVETARIQDILTNTKVERSDPMGRPSATSILAYVIFTSGSTG 10257
QY 91 ----- 90
Db 10258 RPKGVMEHNRNLRVQSNVTSQLPQDLRMAHISNLAFDASIEWEIFTAILNGGALICID 10317
QY 91 ----- 90
Db 10318 YFTLLDSQALRTTFEKARVNATLFAPALLKECLNAPTFLFEDLVLYIGDRLDATDAK 10377
QY 91 ----- 90
Db 10378 IQALVGTVYNAYGPTENTVMSTIYRLTDGESYANGVPIGNAVSSSGAYIMDKORLVPP 10437
QY 91 ----- 90
Db 10438 GYMGELVVGDLARGYTNSTLNADR FVDIVINDOKARAYRTGDRTRVPRKDGSIIEPGR 10497
QY 91 ----- 90
Db 10498 MDQOVKIRGHRVEPAEVEQAMLGNAIHDAAVVVQAVDGOETEMIGFVSMASDRFSEGE 10557
QY 91 -----DDT----- 93
Db 10558 EITNQVQEWEDHFESTAYAGIEADQATLGRDFTSWYNGNLIIDKAEMBEWLDITMQS 10617
QY 94 ----- 93
Db 10618 LLDKEDARPCABIECTGTGMVLFNLPKNDGLESYVIGBPSRSAALFVDKAAQDPGLQGT 10677
QY 94 ----- 93
Db 10678 QILVGTAEIDIKLVKDFHPDVVVVINSVAQYPPSRSYLVQIASELIHMTSVKTIFFGDMRSW 10737
QY 94 -----DP----- 95
Db 10738 ATNRDFLVSRALYTLGDKATKQIQROEVARLEENEDELLVDPAFFTSLTSQWPKVKHVE 10797
QY 96 ----- 95
```

```
Db 10798 ILPKMRMRTSNELSYRYAAVLHICRDGEGRNRYGRRVHVSVEENAWIDPASSGMDRHALVQ 10857
Qy 96 ----- 95
Db 10858 MLDERRDAKTVAIGNIPHSNTINERHFTTSLDTEGEGIAQSLDGSAAWQSAKMAAARCP 10917
Qy 96 ----- 95
Db 10918 CLSVTELVEIGQAAGFRVEVSWARQSOHGALDVVFHHELEDDRVGRVLINFPPTDFERLPP 10977
Qy 96 ----- 95
Db 10978 STGLTSRPLQRIORRRFESQIREQOTLLPPYMPVPSRIVVLERMPLNANSKVORKELARK 11037
Qy 96 ----- 97
Db 11038 ARTLOTIKPSATRVAPRNDIEAVLCDEFQAVLGVTGVMONFFELGCHSLMATKLAARLS 11097
Qy 98 ----- 101
Db 11098 RRLDTRVSKDIFNOPILQDLADVVTGSGAPHEAIPSTPYSGPVQESFSGRLWFLDQLN 11157
Qy 102 ----- 101
Db 11158 LNASWYHMLASRLGRPLRIEALQALATIEARHESLRTTFEBQDGPVQIVRAARNKQL 11217
Qy 102 ----- 101
Db 11218 RIIDVSGTERAYLAALKQBQDAAPDLTAEPGRVALLURLGPDHDLVLSIVMHHIISDGWSV 11277
Qy 102 ----- 103
Db 11278 DILROELGQLYSNASSQAPLPIQYRDFAIWQKDSQIAEHQKQLNWKRLVNSKPAEL 11337
Qy 104 ----- 103
Db 11338 LADTRPKALSGDADVIPDIDDOVYQNLRSFCRRHVTSFVALLAAFRAAHYRLTGAED 11397
Qy 104 ----- 103
Db 11398 ATIGSPIANRRPELEGLICFVNTQCLRIPVKSEDTFTLVKQARETATEAODNODVPF 11457
Qy 104 ----- 111
Db 11458 ERIVSSMVASSRDTSRNPLVQVMFAVHSQHDGLGNIRLEGVEGPKVMAASTRFDAMHFL 11517
Qy 112 ----- 111
Db 11518 EDQMLGNNVVFSKDLFESETIRSVAVFOETLRRGLANPHANLATPLTDGLPSLRSIC 11577
Qy 112 ----- 113
Db 11578 LQVNPQDYPDASVIDVFREQVASIPKSTAVIDAASSQLTYTELDERSSOLATWLRQVTV 11637
Qy 114 ----- 118
Db 11638 PEELVGLAPRSCETIIAFLGIKANLAYPLDVNAPAGRIETILSSLPGNRLILGSDT 11697
Qy 119 ----- 119
Db 11698 QAVKLHANSVRFTTRISDALVESGSPTEELSTRPTAQSLAYYVFTSGTGVPKGMVMEHR 11757
Qy 120 GVT----- 122
Db 11758 GITRLKVNNVAKQPAAMAAIAHLSNIADFASWEIYAPLLNGGTVCIDYTTIDIKAL 11817
Qy 123 ----- 122
Db 11818 EAVFKQHIRGAMLPPALLKQCLVSAPTMISSLEILFAAGDRLSSQDAILARRAVGSGVY 11877
Qy 123 ----- 122
Db 11878 NAYGPTENTVLSTIHNIGENAFSGVPIGNVNSGAFVMDONQOLVSAGVIGELVVTG 11937
Qy 123 ----- 122
Db 11938 DGLARGYTSKLURVDRFIYITLDGNRVRAYRTGDRVRRHRPKDGOIEFFGMRDQOIKIRGH 11997
Qy 123 ----- 122
Db 11998 RIEPAVEQALARDPAISDSAVITQLTDEEPELVAFFSLKGNANGTNGVNGVSDOEKID 12057
Qy 123 ----- 122
Db 12058 GDEQHALLMENKIRHNLOALLPTYMIPSRIIHVDQLPVNANGKIDRNELAVRAQATPRTS 12117
Qy 123 ----- 122
Db 12118 SVSTVYAPRNDIETIICKEFADILSVRVGITDNFFDLGGHSLIATKLAARLSRRLDTRVS 12177
Qy 123 ----- 122
Db 12178 VRDVPDTPVVGOLAASIOOGSTPHEAIPALSHSGPVQOSPAQGRLWFLDRFNLNAWYIM 12237
Qy 123 ----- 122
Db 12238 PFGVRLRGPLRVDAQTALALEERHELLRTTFEBQDVGVMQIVHSPRMDICVVVDISGA 12297
Qy 123 ----- 122
Db 12298 NEDLAKLKEEQOAPPNLSTEVAMRVALPKAGENHHILSIVMHHIISDGWSVDIIFOEBAQ 12357
Qy 123 ----- 122
Db 12358 FYSVAVRGHDPLSQVKPLPIHYRDPFAVWQRODKQAVHVESOLQYWIEQLADSTPAEILSD 12417
Qy 123 ----- 122
Db 12418 FNRPEVLSGEAGTVPIVIEDEVYKLSLFCRNHQVTSFVLLAARVAHRYLRTGAEDATI 12477
Qy 123 ----- 122
Db 12478 GTPIANRNRPELEDLIGFFVNTQCMRIABEHDNPLSVVRVRSTAASAFENQDVPFERL 12537
Qy 123 ----- 122
Db 12538 VSALLPGSRDASRNPLVOLMFVHVSQRNLKQLEGEGETPYTATTTRPDVEFHLPEQD 12597
Qy 123 ----- 122
Db 12598 KGLAGNVVFAADLFEAATIRSVVEVFHEILRRGLQDPDIAISTMPLVDGLAALNSRLPA 12657
Qy 123 ----- 122
Db 12658 VEDIEPDPAEASVVDVFOQVVPANPDALAVTDTSTKLYAELOQSDHVAAMLKSKOKLP 12717
Qy 123 ----- 122
Db 12718 AESIVVVLAPRSSETIVACIGILKANLAYLPMDNSVPEARQAILSEIPGKFVLLGAGV 12777
Qy 123 ----- 122
Db 12778 PIPDNKTADVMMVFISDIVASKTDKSYSGTPRPSASSLAYVIFTSGSTGRPKGMVMEHRG 12837
Qy 123 ----- 122
Db 12838 VISLVKQNASRIPOSLRMAHVSNLAPDASWEIFTTLLNGGTLCISYFTVLDKALSAA 12897
Qy 123 ----- 122
Db 12898 PSDHRINITLLPALLKQCLADAPSVLSLESLYIGGDRLDGADATKVKDLVKGNAYNAY 12957
Qy 123 ----- 122
Db 12958 GPTENSVMSTIYTIETHEFTFANGVPIGTSLGPKSKAYIMDQDQOLVPAGVMGELVYAGDGL 13017
```



```

QY 123 ----- 122
Db 13018 ARGYTPSLNTRGFIHITIDCKQVAYRTGDRVRYRPRDYQIEFFGRLDQOIKIRGHRIE 13077
QY 123 ----- 122
Db 13078 PAVEQALLSDSSINDAUVVSAQNKEGLEMGVYITTOAAQSDVKKEASNKVQWEAHFDS 13137
QY 123 ----- 122
Db 13138 TAVANIGGIDRDALGODFLSWTSMYDGLIPREMOEWLNDTWRSLLDNQPPGKVLEIGT 13197
QY 123 ----- 122
Db 13198 GTGMVFNLGKVEGLQSYAGLEPSRVTAWNKAIETFPFSLAGSARVHVGTAEDISSIDG 13257
QY 123 ----- 122
Db 13258 LRSDLVINSVAQYFPSREYLAELTANLIRLPGVKRIFFGDMRTYATNKDFLVARAVHTL 13317
QY 123 ----- 122
Db 13318 GSNASKAMVROQVAKLEDDEEELLVDPAFFTSLSDOFPDEIKHVEILPKRMAATNELSSY 13377
QY 123 ----- 122
Db 13378 RYAAVIHVGHQMPNGEDEDKQAVKDINPKAWVPFAGTRMDRQALLQLLQDRQRDDVV 13437
QY 123 ----- 122
Db 13438 AVSNIPYSKTIMERHLSQSLDDEDEGTSAVDGTAMISRTQSAKECPALSVADLIEIGK 13497
QY 123 ----- 122
Db 13498 IGFEVEASWARQSORGLDAVHRFEPHSHGVNFRPTEHKGRSSSLNRPPLHLLO 13557
QY 123 ----- 122
Db 13558 SRLEAKVRERLQSLPPYMIPSRITLLDQMLPTNGKVDRKKLARQARVIPRSAASTLD 13617
QY 123 ----- 126
Db 13618 FVAPRTEIENVLCEBFTDLLGVKVGITDNFFELGSHSLATKLSARLSRRLDAGITVKQV 13677
QY 127 ----- 126
Db 13678 FDQVPLADLAASILQSSSRHRSIPSLPYEGPVEQSFQAGRLWFLDQFNIDALWVLPAL 13737
QY 127 ----- 126
Db 13738 RMRGPLOVDALAAALVALVEERHESLRTTTEERDGVGIQVQPLRTTKDIRIIDVSGMRDD 13797
QY 127 ----- 132
Db 13798 DAYLEPQEQTPFDLASEFGWRVALLKLGKDDHILLSVMHHIISDGNSTEVLORELQ 13857
QY 133 ----- 132
Db 13858 FYLAASKGAPLSQVAPLPIQYRDFAVWQOEQVAESORQDYWKQLADSPAEILLAD 13917
QY 133 ----- 132
Db 13918 YTRPNVLSGEAGSVSFVINDSVKLSVFCRSRQVTTFTLLAFAAHYRMTGSDDATI 13977
QY 133 ----- 136
Db 13978 GTPANRRPELENLIGCFVNTQCMRITIGDETFSVLQVQVRSTTATAPENQVPPERI 14037
QY 137 ----- 144
Db 14038 VSTLSAGSRDTSRNPVLQVLLFAVHSQQGLGRIQDGVVDEPVLSTVSTRFDLEPHAFQEA 14097
QY 145 ----- 144

```

```

Db 14098 DRLNGSVMFATDLPQETIQSGFVAVVEVLQGLEQSPQSPQPIATPLAEGIAQLRDAGALO 14157
QY 145 ----- 144
Db 14158 MPKSDYPRNASLVDVFOQOAMASPVSTVAVTDSKLTVAELDRSLDQAASYLRROOLPAE 14217
QY 145 ----- 144
Db 14218 TWAVLAPRSCETIIAFLAILKANLAYMPLDVNTPTSARMEAIISVPGRRLLILVSGVRH 14277
QY 145 ----- 144
Db 14278 ADINPVNAKTMILISDTVTGTDAIGTPPEPLVVRPSATSLAYVIFTSGTGKPKGVNVEHRA 14337
QY 145 ----- 144
Db 14338 IMRLVKDSNVVTHMPPATRMARMAHVNTIAFDVSLFEMCATLLNGGTLVCIDYLLDSTMLR 14397
QY 145 ----- 144
Db 14398 ETFEREQVRAAIFPPALLRQCLVNMPPDAIGMLEAVYVAGDRFHSRDARATQALAGPRVYN 14457
QY 145 ----- 144
Db 14458 AYGPTENAILSTIYNIDKHDPYNGVPIGSAVNSGAYVMDRNOQLLPPGVNGELVVTGE 14517
QY 145 ----- 144
Db 14518 GVARGYTDASLDTDRFVTVTIDGORARAYRTGDRVRYRPGFQIEFFGRLLDQOAKIRGHR 14577
QY 145 ----- 144
Db 14578 VELGEVEHALLSENSVTDAAVVLRTMEEEDPOLVAFVTTDHEYRSGSSNEEDPYATQAA 14637
QY 145 ----- 144
Db 14638 GDMRKRLRSLPYVMVPSRVITLROMPLNANGKVDKDLARRAQMTPTASSSGPVHVAPR 14697
QY 145 -----GVT 147
Db 14698 NETEAAICDEFETILGVKVGIT 14719

```

RESULT 2

```

US-09-369-364A-7
; Sequence 7, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
US-09-369-364A-7

```

Query Match 30.7%; Score 237; DB 4; Length 997;

Best Local Similarity 18.2%; Pred. No. 0.00038;

Matches 61; Conservative 24; Mismatches 62; Indels 180; Gaps 6;

QY 1 AAGGILHLEL-----LVAVG----- 15

Db 53 AGGSFLSYELWPRALKRQDVSRDAPAFYLOYRGRFLRFLNTANQHLLAPCFVSETRR 112

QY 16 -----PD-----VFOAHQED----- 25

```

Db 113 RGLGRAHRAHTPACHLLGEVDPELEGLAAISACDGLKGVFQLSNEDYFIPLDSAP 172
Qy 26 -----|||----- 25
Db 173 ARPGBAHPVVKRQAPERLAQRGDSSAPSTCGVQVPELESRRERWEQROQWRRPRLRR 232
Qy 26 -----TERYVLTNLTNIGAEALLRDPSPSLGAQFRVHLV 55
Db 233 LHORSVSEKWCETLVADAKMVEYHGOPOVESYVLTIMNVAGLFHDPSIGNPIHITIV 292
Qy 56 KXVILTEPEGAPNITANITSSLLSYCGMSQINPEDDTPDGHADLVLYITRFDLELPGDN 115
Db 293 RLVLLEDEEDLKITHADNTLKSFCWKQKSNKMGDAHPLHHTAILLTRKDL-CAAMN 351
Qy 116 R-QVRGVTOLGGACSPWSCLTFTEDTGFDLGVTI 148
Db 352 RPECETLGLSHVAGMCQPHRSCSINEDTGLPLAFTV 386

```

RESULT 3

```

US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin.
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-6

```

```

Query Match 30.4%; Score 235; DB 4; Length 11877;
Best Local Similarity 1.1%; Pred. No. 0.33;
Matches 91; Conservative 21; Mismatches 36; Indels 8329; Gaps 25;

Qy 1 AAGG-----4
Db 450 AAGQAVTDVPADRNAGDFYDPRSAPGRNSRWGFIEDVDRFDAAPFGISPREAEM 509
Qy 5 -----4
Db 510 DPORLALELGEALERAGIDPSSLTGTTRGVFAGAIMDDYATLKHROGGAITPHTVTG 569
Qy 5 -----4
Db 570 LHRGIIANRLSYTLGLRGPMMVDSQSSSLVAHVILACESLRGESELALAGVSLNLP 629
Qy 5 -----4
Db 630 DSIIGASKFGLSPDGRAYTFDARANGYVRGEGGFVWLKRLSRAVADGDPVLAVIRGSA 689
Qy 5 -----4
Db 690 VNNGGAAQMTTPDAQOAEVLEAREHAHERAGTAPADRVYVELHGTGTGTPGDPIEAAALGAA 749
Qy 5 -----ILHLE-----9
Db 750 LGTRPAGQPLLVGSKVTNIGHLEGAAGIAGLIKAVLAVRGALPASLNYETPNPAIPFE 809
Qy 10 -----9
Db 810 ELMNLVNTPELPEPHDQRMVGVSSFGMGGTNAHVLEAPGGCGRGASVVESTVGGS 869

```

```

Qy 10 -----9
Db 870 AVGGVVPVWVSAKSAALDAQIERLAAPASRDRTDGVDAVDAVDAVAVVLGAG 929
Qy 10 -----LLVAVGPD-----17
Db 930 RQAFEHRAVVGSGPDDLAALAAPEGLVRGVASGVRVAFVFPGGGTOMAGKAEELDS 989
Qy 18 --VFQA-----21
Db 990 SAVFAAAMAECEAALSPYVDSLEAVVRQAPCAPTLERVDVVQPVTFVAVMVSARVWQH 1049
Qy 22 -----21
Db 1050 GVTPOAVVGHSGOIEAAAYVAGALSDDAARVVTLRKSIAAHLAGKGMGLSALSED 1109
Qy 22 -----21
Db 1110 LERLAGFDGLSVAAVNGPTATVVGDPVQIEELARACEADGVRARVIVPDYASHSROVEI 1169
Qy 22 -----21
Db 1170 IESELAEVLAGLSPOAPRVFPFSTLEGAWITEPVLDDGGVWYRNLRHVRGFAPAVETLATD 1229
Qy 22 -----21
Db 1230 EGFTHFEVSAHPVLTMALPGTVTGLATLRDNGQDRLVASAEAWANGLAVDMSPLLP 1289
Qy 22 ----HQED-----TER-----28
Db 1290 SATGHHSDLPTVAFOTERHMLGEIALAPAGEPAVQPAVLTEAAEPAELORDEOLRVIL 1349
Qy 29 -----28
Db 1350 DKVRAQTAQVLGYATGGQIEVDRTFREAGCTSLTGLDLNRINAAFGVRMAPSMIFDPPT 1409
Qy 29 -----28
Db 1410 PEALAEQLLLVVHGEAAANPAGAEPAVAAAGAVDEPVAIVGMACRLPGGVASPEDLWRL 1469
Qy 29 -----28
Db 1470 VAGGDALISEFPQDRGMDVEGLYHPDPEHPGTSYVROGGFTIENVAGFADAAFFGISPREAL 1529
Qy 29 -----YVLT 32
Db 1530 AMDPOORLLLETSSWEAVEDAGIDPTSLRGQVGVFTGAMTHEYGPSLRDGGEGLDGYLLT 1589
Qy 33 N-----LNIGAEALLR-----42
Db 1590 GNTASVMSGRVSYTLGLEGPALTYDTACSSSLVAHLAVQALRKGEVDMALAGGVAVMPT 1649
Qy 43 -----42
Db 1650 PGMEFEFSRORLAGDGRSKAFASADGTSWSEGVGLLVERLSDARRNGHOVLAVVRGS 1709
Qy 43 -----42
Db 1710 ALNODGASNLTPNGPSQORVIRRALADARLTSTDVDVVEAHGTGTRLGDPIEAQALIA 1769
Qy 43 -----42
Db 1770 TYGGRDEQPLRLGSLKSNIGHTQAAAAGVSGVIKVMQMRHGLLTKLTHVDEPSDOIDW 1829
Qy 43 -----DPSLG- 47
Db 1830 SAGAVELLTEAVDWPEKQDGLRRAAVSFGISGTNAHVLEBAPVVEGASVVEPSVGS 1889
Qy 48 -----AQF 50
Db 1890 SAVGGVTPVWVSAKSAALDAQIERLAAPASRDRTDDADAGAVDAVAVHLADGRAQF 1949
Qy 51 -----50

```

```
Db 1950 EHRAVALGAGADLVQALADPDGLIRGTASGVGRVAFVPPGGTQWAGMGAEILLSSAVF 2009
Qy 51 -----RVHLV----- 55
Db 2010 AAAMECEAALSPYVDWSLEAVVROAPGAPTLEKRVQVPTFAVMVSLARVWQHGVTP 2069
Qy 56 ----- 55
Db 2070 QAVVGHSGEIAAAVAGALPLDDAARVVTLRSKSIAAHLGKGGKMSLALNEDAVLERL 2129
Qy 56 ----- 55
Db 2130 SDFDGLSVAAVNGPTATVSGDPVQIEELAQACKAGFRARIIPVDYASHSRQVEIIESE 2189
Qy 56 ----- 55
Db 2190 LAQVLAGLSQPQAPRVFFSTLEGTWITEPVLDTGYWYRNLHRVGFAPAIETLAVDEGFT 2249
Qy 56 ----- 55
Db 2250 HFVEVSAHPVLTMTLPETVGLTLLRREGGQGERLVTSIAEAWVNGLPVAMTSLLPATAS 2309
Qy 56 ----- 55
Db 2310 RPGLTYAFOAERYWLENTPAALATGDDWRYRIDWKRLPAEAGSERTGLSGRWLAVTPED 2369
Qy 56 ----- 55
Db 2370 HSAQAAVLTALVDGAKVEVLTAGDDREALARLTALTGTGGFTGVVSLDGLVPQV 2429
Qy 56 -----KMVILTEPE----- 64
Db 2430 AWVQALGDAGIKAPLWSVTQGAHSVGRLDTPADPDRAMLWGLGRVVALEHFERWAGLVDL 2489
Qy 65 ----- 64
Db 2490 PAQPDAAALAHVLTALSGATGEDQTAIRTTGLHARRLARAPLHGRRPRTRDQPHGTVLIT 2549
Qy 65 -----GAPNITANLTSS----- 76
Db 2550 CGTGALGSHAARWMAHGAEHLLVSRSGEQAPGATQTALERTASGARVTTAACOVADPH 2609
Qy 77 ----- 76
Db 2610 AMRTLDDAI PAETPLTA VVHTAGALDDGIVDTLTAEQVRRRAHRAKAVCASVLDLDRDLD 2669
Qy 77 ----- 76
Db 2670 LDAFVLFSSVSTLGI PQGNYAPHNAYLDALAAARRRATGRSAVSAMGPMWGGMAAGD 2729
Qy 77 -----LLSVCGWSOTI----- 87
Db 2730 GVAERLRNHGVPMDPPELALAEALSALGRDETATVADIDWDRFYLAYSGRPPQPLVEEL 2789
Qy 88 ----- 87
Db 2790 PEVRRIIDARDSATSGGSSAOGANPLAERLAAAAPCERTEILLGLVRAQAAVLRMRS 2849
Qy 88 ----- 87
Db 2850 PEDVAADRAFKDIGDSLAVELNRNLT RATGLQLPATLPDPHPTPLALVSLRSEFLGD 2909
Qy 88 -----NPED----- 91
Db 2910 EETADARRSAALPATVAGAGAGAGACTDADDPIAIVAMSCRYPGDIRSPEDLWRMLSEGG 2969
Qy 92 -----DTDP----- 95
Db 2970 EGITFPFTRCWDLDGLYDADPDALGRAYVREGGLHDAAEFDAEFFGVSPREALANDPQ 3029
Qy 96 ----- 95
```

```
Db 3030 QRMILLTSWEAFERAGIEPASLRSGSTGCVFIGLSQDYAARVPNAPRGVEGYLLTGSTPS 3089
Qy 96 ----- 95
Db 3090 VASGRIAYTFGLEGPATTVDTCSSSLTALHLAVRALRSGECTMALAGVAMMATPHMPV 3149
Qy 96 ----- 95
Db 3150 EFSQRALAPDGRSKAFSADADGFGAAEGVGLLLVERLSDARRNGHPVLA VVRGTA VNO D 3209
Qy 96 ----- 95
Db 3210 GASNGLTAPNGFSQQRVIRQALADARLAPGDIDAVETHGTGTSGLDPIEAOOLOATYGKE 3269
Qy 96 ----- 95
Db 3270 RPAERPLAIGSVKSNIGHTQAAAGAAGIIMVLA MRHGTLPKTLHADEPSPHVDWANSGL 3329
Qy 96 ----- 95
Db 3330 ALVTEPIDWPAGTGPRAAVSSFGISGTNAHVVLEQAEPDAAGEVLGADEVPEVSETVAMA 3389
Qy 96 ----- 95
Db 3390 GTAGTSEVAEGSEASEAPAPGSREASLPGLHPMVLSAKDRQSLRGQAAALHAWLSEPA 3449
Qy 96 ----- 95
Db 3450 DLSADGPARLURDVGTYLTATSR TAFAHRAAVTAADROGFLDGLATLAOGGTSAHVHLDTA 3509
Qy 96 -----GHADL----- 100
Db 3510 RDGTTAFLTGTGQSQRPGAGRELYDRHPV FARALDEICAHL DGHLELPLLDVMFAEGSA 3569
Qy 101 ----- 100
Db 3570 EAALLDETRYTOCALFALEVALFRLVBSWGM RPAALLGHSVGETAAAAHVAGVFLSADAAR 3629
Qy 101 ----- 100
Db 3630 LVAARGRLMOELPAGGAMLA VQAAEDEIRVWLETEERYAGRLOVAAVNGPEAAVLSGDAD 3689
Qy 101 ----- 100
Db 3690 AAREABAYMGLGRRTRALRVSHAFSAHMDGMLDGFRAVLETVFERRRPSLTVVSNVTGL 3749
Qy 101 ----- 100
Db 3750 AAGPDDLCDPEYVWRHVHRTVRF LDGVRVLRDLGVRTCLELGPDPGLVTAMAADGLADTPA 3809
Qy 101 ----- 100
Db 3810 DSAAGSPVGPAGSPADSAAGALRPRPLL VALLRRKRSETETVADALGRAHAHGTGPDWH 3869
Qy 101 ----- 100
Db 3870 AWFAGSAHRVDLPTYSFRDRYWLDA PAADTAVDTAGLGLGTADHP LLLGAVVSLPDRDG 3929
Qy 101 ----- 100
Db 3930 LLITGRLSLRTHPWLADHAVLGSVLLPGA MWVLA AHAESA GLRDVRELTLLLEPLVLPE 3989
Qy 101 ----- 100
Db 3990 HGOVELRVTVGAPAGEPGGESAGD GARPVSLHSRLADAPAGTAWSCHATGLLATDRPELP 4049
Qy 101 ----- 100
Db 4050 VAPDRAAMPPQGAEEVPLDGLYERLDGNGLA FGPLFOGLNAVVRVEGEVPADIALPATT 4109
Qy 101 ----- 100
Db 4110 NATAPATANGGSGSAAAPYGIHPALLDASLHAITAVGGLVDEPELVRVPFHWSGVTVHAAG 4169
```

Qy	101	-----	100
Db	4170	AAAARVRLASAGTDAVSLTGDGRPLVSVRLTLRPVTADQAAASRVGGLMHRVAWRP	4229
Qy	101	-----	100
Db	4230	YALASSGEODPHATSYGPTAVLGKDELKVAALLESAGVEVGLYPDLAALSQDVAAAGAPAP	4289
Qy	101	-----VLYITR-----	106
Db	4290	RTVLAPLPAGPADGGAGVGRTVARTLELLQWLADENHLAGTRLLLVTRGAVRDPESGA	4349
Qy	107	-----	106
Db	4350	DGGEDLSHAAANGLVRTAOTENPGRFGLLDLADASSYRTLPSVLSDAGLRDEPOLALH	4409
Qy	107	-----	106
Db	4410	DGTIRLARLASVRPETGTAAPALAPECTVLLTGTGGLGLVARHVVGWVRRLLLVSR	4469
Qy	107	-----	106
Db	4470	RGTAPGABELVHELEALGADVSVAACDVADREALTAVLDAIPAHEHPLTAVVHTAGVLS	4529
Qy	107	-----	106
Db	4530	GTLPSTMTEDVEHVLRPKVDAAFLDLBELTSTPAYDLAAFMFSSAAAVFGAGQAYAAA	4589
Qy	107	-----	106
Db	4590	NATLDALWRRRAAGLPALSLGHLGWLHWAETSGMTGELQADLRMRKSRAGIGGISDAEGIAL	4649
Qy	107	-----	106
Db	4650	LDAALRRDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRARPSAASATTA	4709
Qy	107	-----	106
Db	4710	GTAAGTGTADGAAETAATVLAADRAATVDGPAPORLRLLEFVVGVEVAEVLGHARGHRIDAER	4769
Qy	107	-----FDELPDG-----	114
Db	4770	GFLLDGFDSLTAVELNRNLSAGGLALPATLVFDHPSPAALASHDLDAELPRGASDQDGAG	4829
Qy	115	-----	114
Db	4830	NRNGNENGTTASRSTAETDALLAQLTRLEGALVLTGLSDAPGSEEVLEHLRLSLRSWVTGE	4889
Qy	115	-----	114
Db	4890	TGTGTASGAPDGAGCAEDRPHWAGDGGSGEDGACVPDFMNASAEELFGLLDQDPSTD	4949
Qy	115	-----	114
Db	4950	MSTVNEEKYLDYLRATADLHEARGRLRELEAKAGEPVAIVGMACRLPGVASPEDLWRL	5009
Qy	115	-----	114
Db	5010	VAGGEDAISEFPQDRGWDVEGLYDPNPEATGKSYAREAGFLYEAGEFDADFFGISPREAL	5069
Qy	115	-----	114
Db	5070	AMDPOORLLLESWEAFEHAGIPAAATAGTSVGVFTGVMYHDYATRLTDVPEGIGYLG	5129
Qy	115	-----	114
Db	5130	GNSGSVASGRVAYTLGLEGPVAVTDACSSSLVALHLVAQALRKGEVDMALAGGVTVMST	5189
Qy	115	-----NRQ-----VRG- 120	120
Db	5190	PSTFVEFSRQRLAPDGRSKSPFSSTADGTSWSEGVGVLLVERLSDARRKGRHILAVVRGT	5249
Qy	121	-----	120
Db	5250	AVNQDASSGLTAPNGPSQORVIRRALADARLTTSDDVDVVEAHGTGTRLGDPLEAQAVTA	5309
Qy	121	-----	120
Db	5310	TYGQRDGEOPURLGSLKSNIGHTOAAAAGVSGVIKVMQMRHGVLPLKTLHVEKPTDQVDM	5369
Qy	121	-----	120
Db	5370	SAGAVELLTEAMDWPKDGGCLRRRAAVSFGVSGTNAHVVLEEAPAASETPASEATPAVE	5429
Qy	121	-----	120
Db	5430	PSVGAGLVPLVSAKTPAALDAQIGRLAAFAFASQGRDAAADPGAVARVLGGRABEHRAV	5489
Qy	121	-----	120
Db	5490	VLGTQDDPFAQALTAPEGLIRGTPSDVGRVAFVFPQOGTQWAGMGAELLDVSKEPAAAMA	5549
Qy	121	-----	120
Db	5550	ECESALSRVDSLEAVVRQAPGAPTLERVVDVQVPTFVAVMUSLAKVMQHHGVTPQAVVG	5609
Qy	121	-----	120
Db	5610	HSQGEIAAAVYVAGALTLDAAARVVTLRSKSIAAHLGKGMISLALSEATRQRIENLHG	5669
Qy	121	-----	120
Db	5670	LSIAAVNGPTATVVGSDPTQIQLAQACEADGVRRARIIPVDYASHSAHVETIESELAEVL	5729
Qy	121	-----	120
Db	5730	AGLSPTPEVFPFFSTLEGAMITEPVDGTYWYRNLRRHVRGPAVETLATOGETHFTEV	5789
Qy	121	-----VTOLG-----	125
Db	5790	SAHPVLTWLTBETVTGLGLTLLRREGQGOERLVTSLAEAWNTGLTDWAPVLTATGHHPEL	5849
Qy	126	-----	125
Db	5850	PTYAFORRHVHDSPAVOGSQDSWRYRIDWKLAVADASERAGLSGRWLVVVPEDRSA	5909
Qy	126	-----	125
Db	5910	EAPVLAALSGAGADPVQDVSPLGDRQLAATLGEALAAAGGAVDGVLSLLAWDESAHP	5969
Qy	126	-----	125
Db	5970	GHPAPFTRGTGATLTLVQALEDAGVAAAPLWCVTHGAVSVGRADHVTSPAQAMVWGMGRVA	6029
Qy	126	-----G 126	126
Db	6030	ALERPERWGGLIDLPSDADRAALDRMTTVLAGGTGEDQVAVRASGLLARLVRASLPAGH	6089
Qy	127	ACSPTW-----	132
Db	6090	TASPMWQADGTVLVTGAEPAAEAARRLARDGAGHLLHLLHTTSPGSEAGCTSGAEDSG	6149
Qy	133	-----	132
Db	6150	LAGLVASLADLGATATVVTCDLTDAEAAARLLAGVSDAHPLSAVLHLPPTVVDSEPLAATD	6209
Qy	133	-----	132
Db	6210	ADALARVVTAKAALHLDRLLREAAAAAGRPVVLVLFSSVAALWGGAGOGAYAAGTAFGL	6269
Qy	133	-----	132
Db	6270	DALAGHRADGPTVTSVANSWPWEGSRVTEGATGERLRLGLRPLAPATALTALDTALGHG	6329
Qy	133	-----	132

Db 6330 DTAVTIADVMSFAPGFTTARPGTLLADLPEARRALDEQQSTTAADTVLSRELGALTG 6389
Qy 133 ----- 132
Db 6390 AEOQRKQELVREHLAVLNHPSPBAVDTCRAPRDLGFDLSLTAVELNRNLKNATGUALPA 6449
Qy 133 ----- 132
Db 6450 TLVFDYPTPTLAEFLLEILGEQAGAGEQLPVDGGVDDEPVAIVGMACRLPGGVASPED 6509
Qy 133 ----- 132
Db 6510 LMRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYCRAGGFLEAGEFADFEGISP 6569
Qy 133 ----- 132
Db 6570 REALANDPOORLLETSEWEAVEDAGIDPTSLQQQGVGFAGTNGPHYEPDLLRNTAEBLEG 6629
Qy 133 ----- 132
Db 6630 YVGTGNAASIMSGRVSYTLGLEGPAVTVDTCSSSLVALHLAVQALRKGECCGLALAGGVT 6689
Qy 133 ----- 132
Db 6690 VMSTPTTFVFSRQRGLAEDGRSKAPASADGFGPAEYGVMLLVERLSDARRNGHRVLAV 6749
Qy 133 ----- 132
Db 6750 VRGSVNDQASNGLTAPNGPSQORVIRRALADARLTADVDVVEAHGTGTRLGDPIDIAQ 6809
Qy 133 ----- 132
Db 6810 ALIATYGOGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIMQVQMRHGVLPKTLHVDRPSD 6869
Qy 133 ----- 132
Db 6870 QIDWSAGTVELLTEAMDMWRKQEGGLRAAVSSFGISGTNAHIVLEEA PVDEDAPEDEPS 6929
Qy 133 ----- 132
Db 6930 VGVVYVPLVSAKTPAALDAQIGRLAAAFASQGTDAADPGAVARVLGAGRAQFEHRAVALG 6989
Qy 133 ----- 132
Db 6990 TQDDLAALAAPEGLVRGASGVRAVFVPGOGTQWAGMGAELLDVSKFAAAMAECE 7049
Qy 133 ----- 132
Db 7050 AALAPYVDSLEAVVRQAPGAPTLERVDVQPVTFAMVMSLAKVMQHGVTPQAVVGHQS 7109
Qy 133 ----- 132
Db 7110 GEIRAAAYAGALSDDAARVVTLSKSI GAHLA GOGGMLSLSEAAVVERLAGFDGLSV 7169
Qy 133 ----- 132
Db 7170 AAVNGPTATVVSGDPTQIQELAQACEADGVRAIIPVDYAGSAHVETIESELADVLAGL 7229
Qy 133 ----- 132
Db 7230 SPOTPOVPFFSTLEGAWITEPALDGGYWYRNLRHVRGPAPAVETLATDEGTFHEVSAH 7289
Qy 133 ----- 132
Db 7290 PVLTMALPETVTGLTLRRDNGQHRLTTS LAEAWANGLTVDMASLLPTTTHPDLPITYA 7349
Qy 133 ----- SCIT
Db 7350 FOTERYWPQDLSAAGDITSAGLGAHEPLLCAAVALADSDGCLLTGSLRTHPWLADH 7409
Qy 138 ----- 137

Db 7410 AVAGTVLLPGTAPVELAFRAGDOVGCOLVBELTLDAPLVLPRGAVRVOLSVGASDESGR 7469
Qy 138 ----- 137
Db 7470 RTFGLYAHPEDAPOGEAEWTRHATGVLAARDRTAPVADPEAMPPEGAPVDVDGLYERFA 7529
Qy 138 ----- 137
Db 7530 ANGYGYGFLFOGVRGVWRRGDEVFADVALPAEVAGAGARFGLHPALLDAVAQAAGACRG 7589
Qy 138 ----- 137
Db 7590 VRRGHAAVRLERDLLYAVGATALRVRLAPAGPDTVSYSAADSSGQPVFAADSLTVLPVD 7649
Qy 138 ----- 137
Db 7650 PAQLAATSDPTLDALHLEWTAWDGAQAALPGAVVLOGDADGLAALRAGGTEVLSPDDL 7709
Qy 138 ----- 137
Db 7710 TDLVEAVDRGETPAPATVVLVACPAAGPDGPEHVREALHGSLLAQMLADERFTDGRLLV 7769
Qy 138 ----- 137
Db 7770 VTRDAVAARSQDGLRSTGQAAVMGLGRSAQTESPGRFVLLDLAGEARTAGDATAGDLTT 7829
Qy 138 ----- 137
Db 7830 GDATVGTSGDAALGSALATALGSGEPQALRQCALLVPRLARAAAPAAADGLAADGLA 7889
Qy 138 ----- 137
Db 7890 ALPLPAALWRLPEPGTSGLESITAAPGDAETLAPEPLPGQGVRIAIRATGLNFRDVL 7949
Qy 138 ----- 137
Db 7950 ALGMYDPALMGTEGAGVVTATGPGVTHLAPGDRVMGLLSGAYAPVVVADARTVARMEG 8009
Qy 138 ----- 137
Db 8010 WTPAOGASVPVFLTAVYALURDLADVKPGERLLVHSAAGVGMAAOLARHWGVVHGTA 8069
Qy 138 ----- 137
Db 8070 SHGWDALRALGLDDAHIASSRTLDPESAPRAASGGAGMDVWLSLAREFVDASRLG 8129
Qy 138 ----- 137
Db 8130 GGRFVEMGKTDVDRDAERVAADHPGVGYRAFPLGEAGPERIGEMLAEVIALFEDGVLRLP 8189
Qy 138 ----- 137
Db 8190 VTTWVRRARDAPRHVSQARHTOKVVLTMPSGLDPGTVLLTGGTGALGGIVARHVVGW 8249
Qy 138 ----- 137
Db 8250 GVRRLLVSRRTDAPGAGELVHELEALGADVSAACDVADREALTAVLDSIPAHEPLTA 8309
Qy 138 ----- 137
Db 8310 VHTAGVLSGTLPSMTAEDVEHVLRPKVDAAPFLDELSTPGYDLAAVFMFSSAAAVFG 8369
Qy 138 ----- 137
Db 8370 GAGQAYAAANATLDALAWRRRTAGLPALSLGHLWAETSGMTGGSLSDTDRSLARSGAT 8429
Qy 138 ----- 137
Db 8430 PMDSELTLSLLDAAMRRDPALVPIALDVAALRAQORDGMLAPLLSGLTRSGRVGGAPVN 8489
Qy 138 ----- ED 139
Db 8490 QRRAAAGGAGEADTDLGGRLAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRD 8549

QY 140 TGFDL----- 144
Db 8550 TGFDSLTVLNRNLNAATGLRLPATLVDPHPTPGELAGHLDELATAAGGWAEGTSG 8609
QY 145 ----- 144
Db 8610 DTASATDRQTAAELAELEGLVSLAPAAAGRPPELAARLALAAALGGDDGDDATDLDE 8669
QY 145 ----- 144
Db 8670 ASDDLFSFDIKELGSDFMANNEDKLDYLKRVTAELQONTLRLREIEGRTHEPVAIVG 8729
QY 145 ----- 144
Db 8730 MACRLGGVASPEDLMQLVAGDGAISEPQDRGNDVEGLYDPPDASGRTYCRSGFLH 8789
QY 145 ----- 144
Db 8790 DAGEFDADFGISPREALAMDPOORLSLTTTAWEAIESAGIDPTALKGSLGVFVGWHTG 8849
QY 145 ----- 144
Db 8850 YTSQTTAVQSPEGLHVSAAALGFLSGRIAYVLGTGDPALTVDTACSSSLVALHLAVQ 8909
QY 145 -----GVTI 148
Db 8910 ALRKGCDMALAGGTVV 8926

RESULT 4

US-09-036-987A-6
; Sequence 6, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow AgroSciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5588 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-036-987A-6
Query Match 27.9%; Score 216; DB 4; Length 5588;
Best Local Similarity 1.7%; Pred. NO. 0.6;
Matches 92; Conservative 25; Mismatches 29; Indels 5123; Gaps 30;
QY 2 AGGI----- 5
Db 230 AGGISMATPGAFVFSRQALASDGRCKPPADAADGTGCGAGMLLLERLSDARNQH 289
QY 6 ----- 5
Db 290 PVLAAVGSAINQDGTSGNLTAPSGAQORVIRQALANAGLSPAEVDVVVEAHGCTGALGD 349
QY 6 -----LHL 8
Db 350 PIEAQUALIATYGANRSADHPLLLSLKSNIGHTQAAAGVAGVIKSVLAIRHREMPRLHI 409
QY 9 ----- 8
Db 410 DOPSHVMSAGAVRLLTDSVMDPLGRPRRAGVSSFGMSGTNAHLIIVEVSDPEVSGST 469
QY 9 ----- 8
Db 470 EPTGAPWPLSGKTETALREQAAELLVVTEHPGLGDVGYSLATGRAAMEHRAVVAD 529
QY 9 ----- 8
Db 530 DRDSFVAGLTALAAGVPAANVVGAAADCKGVAFVPGQSHWQGMARELSSESPVFRRK 589
QY 9 ----- 8
Db 590 LAECAATAPIYVDWLLGLVLRGDPDAPALDRDDVTQLALFAMVSLAELWRCGVEPAAV 649
QY 9 ----- 8
Db 650 VCHSGEITAAHVAGALSITDAVRIIARCDASVLTGKGMALALPESAVVKRIAGLP 709
QY 9 ELLVAV----- 14
Db 710 ELTVAANGPGSTVVGSEPSALERLQELTAENVOTRRVIGIDYASHSPQIAOVGRLDR 769
QY 15 ----- 14
Db 770 LGEVSGEPAEIAFYSTVTGERTDTGRDLADYVYQNLROPRFOQTVMADOGYRFFVEV 829
QY 15 ----- 14
Db 830 SPHPLLTAGIQETLEAADAGGVVVGSLRRGEGSRRLTSLAECQVRGLPVNWEQVFLNT 889
QY 15 ----- 14
Db 890 GARRVPLPTYPFQORYWLESAEYDAGDLGSLGSLLSAEHPLLGAAVTLADAGGFLLTGKL 949
QY 15 ----- 14
Db 950 SVKTPQLADHVVGAILLPGTAFVEMILIRADQVGCGLIEELSLTTPPLVLPAATGAVQV 1009
QY 15 -----GPD----- 17
Db 1010 IAVGPDGAGRRSVRVHSCRODDAVPDQSWTCHATCTLTSSDHQDAGQDPGIWPPNDAVA 1069
QY 18 -----VFO-----AHQEDT----- 26
Db 1070 VPLDSFYARAAERGFDFGPAFOGLQAAMKRGDEIFAEVGLPTAHREDAGRFGIHPALLDA 1129
QY 27 ----- 26
Db 1130 ALQALGAAEEDPEGWLPFAMQVSLKATGALSRLHLPAGANAVSVPTTDTTGOAVLS 1189
QY 27 ----- 26

```
Db 1190 IDSLVLRQISDKOLAAAAAMEHESLFRVDWKRIISPGAAKPFVSWAVIGNDELARACGSALG 1249
QY 27 ----- 26
Db 1250 TELHPDLTGLADPPDPPVPCASRQDLDVASEARAATQRMLDLIQDWLAAARFAGSRL 1309
QY 27 ----- 35
Db 1310 VVVTGGAAGTGAEGVSDLVHAASWGLLRSAQSENDFRFLVDVDTAESWRALAAAVRS 1369
QY 36 ----- 35
Db 1370 GEPQLALRAGEVRVRLARCVAEADSRI PVPAGADGTVLISGGTGLGLVARHLVAERG 1429
QY 36 ----- 39
Db 1430 RRLVLAGRGRWSAPGVDLVDLGLGAAVEVASCVDGDAQDLRLTTISAEFFLRGVV 1489
QY 40 ----- 39
Db 1490 HAAGALADGWESLTPHVAKVFGPKAAGAWHLHLLTDLDSFFVLFSSFSFGVAGAAGQ 1549
QY 40 ----- 39
Db 1550 GNYAANAFLDGLAQRRTAGLPVAVSLAWGLWEQPSGMTGALDAAGRSRIARTNPPMSAP 1609
QY 40 ----- 43
Db 1610 DGLRUFEMAFRVPGESLLVPVHVLDNALRADAADCGVPALLRDLVPAVRRAVNESADV 1669
QY 44 ----- 51
Db 1670 NGLVRLRLPDLDQETOLLGLVREHVSAGVAVLGHSGAVEGADRAFRLGDFSLSGVEFRN 1729
QY 52 ----- 51
Db 1730 RLGGVLGVRLPATAVFDYPTPRALVRFLLDKLIIGVVEAPTAPAAVAATAVDDPVVIVGM 1789
QY 52 ----- 51
Db 1790 GCRCYGVSSPEELWRLVAGGLDAVAFPPDGRGWDAQFLDPDRLGTSYVCEGGFLRD 1849
QY 52 ----- 51
Db 1850 AAEFDAGPFGISPREALAMPQORLLLEAVMETVERAGIDPLSLRGSRTGVFAGLMHHDY 1909
QY 52 ----- 54
Db 1910 GARFITRAPEGFEGLNGSAGGVFSGRVAYSFGFEGPAVTVDTCSSSLVALHLAGOAL 1969
QY 55 ----- 54
Db 1970 RSGECDALAGGVTVMATPGMFVPSRQRLAADGRCKSFRAAADGTGCGEGAGLVLLER 2029
QY 55 ----- 54
Db 2030 LSDARRNGHAVLAVRGSVAVNQDGSANGLTAPNGPSQORVITQALASAGLSVSDVAVEA 2089
QY 55 ----- 58
Db 2090 HGTGTRLDGPIEAQALIAITYCGGRDSRPLWLGSVKSNIGHTQAAAGVAGVIRKVMAMRH 2149
QY 59 ----- 62
Db 2150 QQLPATLHVDEPTSEVDWSAGDVOLLTENTPMPGNHPRRVGVSSFGISGCTNAHVILEQA 2209
QY 63 ----- 62
Db 2210 SKTPDETADKSGPDSSESTVDLPVPLIVSGRTPAALSQAASALLSYLGERGDI STLDAAF 2269
QY 63 ----- 62
Db 2270 SLASSRAALEBERAVVLGADRETLTLLSGLEALASGREASGVSGSPVSGGVGFVFAGOGGOW 2329
QY 63 ----- 62
Db 2330 LGMGRGLYSVPFVADAPDEACAGLDHAGQDVGVRDVFVFGSDGSLDRTLMAOSGLFAL 2389
QY 63 ----- 65
Db 2390 QVGLLSLLGSGVGRPGVVVLGHSVGEPAAPAAVAGVLSLPDAARMVAGRARLMAQLPSGGM 2449
QY 66 ----- 65
Db 2450 LAVAAGEEQRLPLADRVDGAGIAAVNAPESVLSGDREVLDDIAGALDGGIRWRLRV 2509
QY 66 ----- 65
Db 2510 SHAFHSYRMDPMLQEPFAEIAARSVDYRRGDLPVVSSTLTGELDTAGMATPEYVWVRQREP 2569
QY 66 ----- 65
Db 2570 RFADGVRVLAQGVATIFELGPDATLSALIPDCHSWADQAMPI PMLRKDRTEVTVVA 2629
QY 66 ----- 65
Db 2630 ARAHTRGVVPEVMSAYFAGTGARRVELPTAFQORQYWLETSDYDVTGIGLAAAEHPLLG 2689
QY 66 ----- 65
Db 2690 AVVALADGDMVLTGRLSVGTHPWLAOHRVLGEVVVPGTATILEMALHAGARLGCORVEEL 2749
QY 66 ----- 65
Db 2750 TLETPLVPERAAGAGSRGPAGGTTVSIETAERVRTNDAIEIOLLVNADEGGRVSVL 2809
QY 66 ----- 65
Db 2810 YSRPAGSRGGWTRHATGELVVGTTGTRAVPDMWAEGBESIALDEFYVALAGNGFEYGP 2869
QY 66 ----- 65
Db 2870 LFOQLQAAWRGDEVLAETAPAEADAMASGYLLDPALLDAALQASALGRPEQGGANLP 2929
QY 66 ----- 65
Db 2930 FSFTGVELSAPAGTISRVRLETRRRPDAISVAVMDESGRLASIDSLRLRSSVSQLANRD 2989
QY 66 ----- 65
Db 2990 AVRDALFEVTPVATQSTEPGRWALLGDTACGKDDLIKLATDSADRCADLAALAEKLD 3049
QY 66 ----- 65
Db 3050 SALVPDVVVYCEQADPGTGAAALAEOTOTLALLOAWLAEPRLAEARLVVVTCAAVTTA 3109
QY 66 ----- 65
Db 3110 PSDGASELAHAPLWGLLRAAQVENPGQFVLADVDTAESWRALPSALGSMEPQLALRKGA 3169
QY 66 ----- 65
Db 3170 VRAPRLASVAGQIDVPVAVVADPRTVLIISGGTGLLGAVARHLVTERGVRRRLVLTORRGW 3229
QY 66 ----- 82
Db 3230 DAPGITELVCELNGLCVVDVVACDVADRADLESLLAAVPAEPPLCGVVHAAGALADGVI 3289
QY 83 ----- 84
Db 3290 ELSLPDDVGAVFGPKAAGAWNLHETRLTDLSFFALFSSLSGVAGAPGOGNYAAANAFLD 3349
QY 85 ----- 84
Db 3350 ALAHYRRSOGLPAVSLAWGLWEQPSGMTETLSEVDRSRIARANPPLSTKEGLRFLFAGLA 3409
```

QY 85 ----- 84
Db 3410 LDRAAVPAKLDRTEFLAEQARSGLPALLTALVPPIRNRNRASGTGLADEGTLGVVREH 3469
QY 85 ----- 84
Db 3470 AAVALGYSSAADYGVVERAFDGLGDSLSGVELRNRLAGVLGVRLPATAVFDYPTPRALAR 3529
QY 85 ----- 91
Db 3530 FLHQELADEIATTPAPVTTTRAPVAEDDLVAIVGMGCRFPQGVSSPEELWRLVAGGVDAV 3589
QY 92 ----- 91
Db 3590 ADPPADRGWDLAGLFDPPDPERAKTYVREGAFLTDADRFDAGFFGISPREALAMDPOQRL 3649
QY 92 ----- 91
Db 3650 LLELSWEAIERAGIDPGSLRSGRTGVFAGLMYHDYGARFASRAPEGFEGVLGNGSAGSVA 3709
QY 92 ----- 91
Db 3710 SGRIAYSGFEGPAVTVDTACSSSLVALHLAGQSLRSGECDLALAGGVTVMSPTGFVEF 3769
QY 92 ----- 91
Db 3770 SRORGLAPDGRCKSPAESADGTGWGEGAGLVLLERLSDARRNGHRVLAVVRGSVAVNQDGA 3829
QY 92 ----- DTD 94
Db 3830 SNGLTAPNGSQORVIOQALASAGLSVSDVDVAEAGTGTRGLGDPTEAQALIAITYGRDRD 3889
QY 95 PG-----HAD----- 99
Db 3890 PGRPLMLGSVKSNIIGHTQAAAGVAGVIKMYMAMRHGOLPRTLHVDAPESSQVDSAGRVOL 3949
QY 100 ----- 99
Db 3950 LTENTWPDGRCRCVGVSSFGISGTNAHVILSQSTQGMQQAABDPDSSPVLDPVVPVVV 4009
QY 100 ----- LVLYITR----- 106
Db 4010 SGKTPEALSAQAATLATYLDQNDVSDPLDVGISLAVTRSDALDERAVVLGSDRDTLLSGLN 4069
QY 107 ----- FD----- 108
Db 4070 ALAAGHEAGVVTGPVIGIGRTGVFAGQGGWLMGMRRLYSEFPAPAFACAEALDA 4129
QY 109 ----- 108
Db 4130 NLGREVGVRDVFSGDESLLDRTLMAQSGLFALOVGLWELLGTWGVRRPSVVLGHSVGELA 4189
QY 109 ----- 108
Db 4190 AFAAGVLSNAEAARLVAGRARLMQALPCSGAMLAVSATEARYGPLLDDGVRDVRGVAAYN 4249
QY 109 ----- 108
Db 4250 APGSVVLGSDRDVLDGTAGRLDGGIRSRMLRVSHAPSHRMDPMLAEFAELARSVDYRS 4309
QY 109 ----- 108
Db 4310 PRLPVSTLTGNLDDGVGMATPEYVWRQVREPVRFDAGVOALVDQGVDTTIVELGPDGALS 4369
QY 109 ----- 108
Db 4370 SLVQECVAESGRATGIPLVRRDRDEVRTLDAQAHTRCGAVDWGSGFFAGTRATQVDLP 4429
QY 109 ----- 108
Db 4430 TYAFORORYWLEPSDSDGTGVLGTGAEPHLLGAVVPVAGDGVLLTGRLSVGTHPWLA 4489

RESULT 5

US-09-370-700-6

; Sequence 6, Application US/09370700

; Patent No. 6274350

QY 109 -----LEL----- 111
Db 4490 HRVLGEVVVPGTALLEMAWRAGSOVCERCVEELTLEAPLVLPERGAAAQVQAVGAPDEAG 4549
QY 112 -----PGNRQV-----RG 120
Db 4550 RRSLLQYSGADEGDMRRRIASGLLAQANAVPADSTAMPDPGAGQVDIAEFYERLAERG 4609
QY 121 VT----- 122
Db 4610 LTYGPVFOGLRAWRHGGDIFPAELAGSPDASGFIHPALLDAALHAMALGASPDSEARLP 4669
QY 123 -----OLG----- 125
Db 4670 FSWRGAQLVRAEGAALRVRLSPLGSAVSLTLVDATGRRVAAVESLSTRPVSTDOIGACR 4729
QY 126 -----GACSPTH----- 132
Db 4730 GDQERLLHVEMVRSAESAGMSLTSCAVVGLGEPEHAAKTKTGVOVESHADLASLATEVA 4789
QY 133 ----- 132
Db 4790 KRGSAPGAVIVPCPRPRMOELPTAARRATQAMAMLOQMADDPRFVSTRLLILLTHRAVS 4849
QY 133 ----- 132
Db 4850 AVAGEDVLDLVHAPLWGLVRSQAEPDRFALJDMDDERASQTALAEALTAGEAQLAVRS 4909
QY 133 -----SCLITEDTG----- 141
Db 4910 GVVLPRLQGVKSGEAPRWDEGTVLVTGGTGLGALLARHLVSAHGVRHLLASRRGL 4969
QY 142 ----- 141
Db 4970 AAPGADELVAELEQACADVAVVACDSADRSLSARLVASVPAENPLRVVTHAAAGVLDGVL 5029
QY 142 ----- 141
Db 5030 MSMSPERLDVLRPKVDAAYLHELTRGLSFAVLFSVSVAGLFGAGOSYAAAGNAFLD 5089
QY 142 ----- 141
Db 5090 ALAHCRAOQLPALSLASGLWASIDGMAGDLAAADVRLSRAGIGPLSAPGGLAFDAV 5149
QY 142 ----- 141
Db 5150 GSDEPLLPVRLDVEALRVQARSVQTRIPEMLHGMMGSPSRPTPTTSRVEPLHERLAGUS 5209
QY 142 ----- 141
Db 5210 EGERRQOVLQVRADIAVVLGHGRSSVDIEKPLAELGSDVLTAEIERNRLATATGLRUP 5269
QY 142 ----- 141
Db 5270 ATLAFDHGTAALAHQVCAQLGTATAPAPRRRDDNDATEDPVRSLFQOAVAAGRILDGMDL 5329
QY 142 ----- 141
Db 5330 VKVAAQLRPVFGSPGELESPLKPVQLSRGPBEALVCMPPALICMPPAQOYARIAAGFRDV 5389
QY 142 ----- 141
Db 5390 RDVSVIPMPGFTAGEPLPSAIEAVRVTOAEAVLQEFAGGSFVLVGHSSGGWLAHEVAGEL 5449
QY 142 -----FDLGVT 147
Db 5450 ERRGVVPAGVLLDVTIPGEITPRFSVMAHRTYEKLATFTDMQDVGIT 5498

; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldfon, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 5588
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-6

Query Match 27.9%; Score 216; DB 4; Length 5588;
Best Local Similarity 1.7%; Pred. No. 0.6;
Matches 92; Conservative 25; Mismatches 29; Indels 5123; Gaps 30;

Qy	2	AGGI-----	5
Db	230	AGGISVMTGPAVFESRQRALASDGRCKPFADAADGTGWGEGAGMLLLERLSDRRNH	289
Qy	6	-----	5
Db	290	PVLAAVGSAINQDTSNGLTAPSGPAQQRVIRQALANAGLSPAEVDVVEAHGTGALGD	349
Qy	6	-----LHL 8	
Db	350	PIEAQALITACNRSADHPLLGLSKNIGHTQAAAGVAGVIKSVLAIRHREMPRLHI	409
Qy	9	-----	8
Db	410	DOPSHVDSAGAVRLTDSVDWPDGRPRAGVSSFGMSGTNAHLIVEVSDPVS GST	469
Qy	9	-----	8
Db	470	EPTGAPWPLSKTETALREQAAELLSVVTETHEPEGLDVGYSLATGRAHEHRAVVAD	529
Qy	9	-----	8
Db	530	DRDSFVAGLTALAAGVPAANVVQGAADCKGVAFVFPQGGSHWQGMARELSESSPVFRK	589
Qy	9	-----	8
Db	590	LAECNAATAPYVDHSLGVLGRDPPAPALDRDDVIQLALFAMVSLAELWRSCGVEPAV	649
Qy	9	-----	8
Db	650	VHSGEIAAAHVAGALSITDAVRIIARCDVSAVLTGKGMLALPESAVVVKRIAGLP	709
Qy	9	ELLVAV-----	14
Db	710	ELTAAVNGPGSTTVSGEPSALERLQTELTAENVQTRRVGIDYASHSPQIAQVQGRLLDR	769
Qy	15	-----	14
Db	770	LCEVSEPAETAFTYSTVTGERTDTGRDLADYQNLROPVRFOQTVMARMADQGYRFFVEV	829
Qy	15	-----	14
Db	830	SPHPLLTAGIQTLEAADAGGVVGSLLRRGGSGSRMLTSLAECQVRGLPVNWEQVFLNT	889
Qy	15	-----	14
Db	890	GARRVPLPTYPFORYWLESAEYDAGDLGSLLSAEHPLLGAATLADAGGFLLTGKL	949

Qy	15	-----	14
Db	950	SVKTPWLAHDVVGGAILLPGTAFVEMLIRAAQVQCDLIEELSLTTPVLVLPATGAVQVQ	1009
Qy	15	-----GPD-----	17
Db	1010	IAVGGDEAGRRSVRVHSCRDDAVPQDSWTCHTATGTLTSSDHQDAGQDGIWPPNDAVA	1069
Qy	18	-----VFQ-----AQEDT-----	26
Db	1070	VPLDSFYARAABRGDFGPAFQGLQAANKRGDEIFAEVGLPTAHREDAGREGIHPALLDA	1129
Qy	27	-----	26
Db	1130	ALQALGAABEDPDEGWLPPFAWQGVSLKATGALSRLVHLVPAGANAVSVFTTDTTGOAVLS	1189
Qy	27	-----	26
Db	1190	IDSILVLRQISDKLAAARAMEHESLFRVDMWKRIISPAAKPVSWAVIGNDELARACGSALG	1249
Qy	27	-----	26
Db	1250	TEHPDLTGLADPPDVVVVPCGASRODLVASEARAATORMLDLIQDWLAAARFAGSRL	1309
Qy	27	-----ERYVLTNLN-----	35
Db	1310	VVVTGGAATGPAEGVSDLVHAASWGLLRSAQSENPDPRFVLVDVDTAESWRALAAAVRS	1369
Qy	36	-----	35
Db	1370	GEPOLARAGEVRVPRLCARCVAAEDSRI PVPGADGTVLSGGTGLLGLVARHLVAERGV	1429
Qy	36	-----ICAE-----	39
Db	1430	RRLVAGRRGWSAPGVTDLVDELGLGAAVEASCDVGDRAQLDRLLTTISAEFPLRGV	1489
Qy	40	-----	39
Db	1490	HAAGALADGVVESLTPEHVAKVFGPKAAGAWHLHELTLDDLDFVFLFSSFSGVAGAAGQ	1549
Qy	40	-----	39
Db	1550	GNVAAAAPLUDGLAQHRRTAGLPAVSLAWGLWEOPSGMTGALDAAGRSRIARTNPPMSAP	1609
Qy	40	-----LLRD-----	43
Db	1610	DGLRLFEMAFRVPGESLLVPVHVDLNALRADAADGGVPALLRLDLVPAPVRRSAVNESADV	1669
Qy	44	-----PSL-----GAQFR-----	51
Db	1670	NGLVGLRLRPDLDDQETQLLCLVREHVSAVLGHSGAVEGACDRAFRDLGDFSLSGVBEFRN	1729
Qy	52	-----	51
Db	1730	RLGGVLGVRLPATAVFDYPTPRALVRFLDKLIGVEAPTPAAPAAVAATDDPVIWGM	1789
Qy	52	-----	51
Db	1790	GCRYPGVSSPEELMRLVAGGLDAVAEFPDDRGWDQAGLFPDPPDLGTSYVCEGFLRD	1849
Qy	52	-----	51
Db	1850	AAEFDAFGFICISPREALAMDPOORLLLEVANWETVERAGIDPLSLRGSRCTGVFAGLMHHDY	1909
Qy	52	-----VHL-----	54
Db	1910	GARFITRAPBGPBGVYLGNGSAGGVFSGRVAYSFGPEGPAVTVDTCSSSLVALHLAQAL	1969
Qy	55	-----	54
Db	1970	RSGECDLALAGGVTVNMTGPMFVFSRORGLAADGRCKFAAAADGTGMGEGAGLVLLER	2029

QY 55 ----- 54
Db 2030 LSDARRNGHAVLAVVRGSAVNODGASNGLTAPNGPSQORVITOALASAGLSVSDVAVEA 2089
QY 55 -----VKMV----- 58
Db 2090 HGTGTRLDPIEAQALATYCGOGRSDRPLMLGSKVSNIGHTQAAAGVAGVIRKMMAMRH 2149
QY 59 -----ILTE----- 62
Db 2150 GOLPATLHVDEPTSEVDWSAGDVOLLTENTPMPGNHPRRVGVSSFGISGTNAHVILEQA 2209
QY 63 ----- 62
Db 2210 SKTPDETADKSGPSESTVDLPVPLIVSGRTPAALSQAQALLSYLGERDITLDAAF 2269
QY 63 ----- 62
Db 2270 SLASSRAALEERAVVLGADRETLISGLEALASGREASGVSGSPVSGGVGFVAGQGQW 2329
QY 63 ----- 62
Db 2330 LMGRLYVPFPFADAFDEACAGLDAHLGQDVGVDRVDVFGSDGLDRTLWAQSGLFAL 2389
QY 63 -----PEG----- 65
Db 2390 QVGLLSLLGSGVVRPGVVLGHSVGEFAAANAAGVLSLPDAARMVAGRARLMQALPSGAM 2449
QY 66 ----- 65
Db 2450 LAVAAGEEOLRPLIADRVDGAGIAAVNAPESVVLSGDREVDDIAGALDQGIWRRLRV 2509
QY 66 ----- 65
Db 2510 SHAFSYRMDPMLQEFAPAIARSVDYRRGDLPVVSTLTGELDTAGVMATPEYVWRQVREP 2569
QY 66 ----- 65
Db 2570 RFADGVRVLAQGVATIFELGPDATLSALIPDCHSNDAQMPIMLRKORTETETVVAAV 2629
QY 66 ----- 65
Db 2630 ARAHTRGVPEWSAYFAGTGARRVELPTYAFQRYWLETSDYGDVTGIGLAAAHEPLIG 2689
QY 66 ----- 65
Db 2690 AVVALADGCMVLTGRLSVGTHPWLAQHRVLGEVVVPGTALILEMALHAGARLCCORVEL 2749
QY 66 ----- 65
Db 2750 TLETPLVPERAAGAGSRGPAGGTVSIETAERVRTNDAIEIQLLVNAPDEGRRRVSL 2809
QY 66 ----- 65
Db 2810 YSRPAGSRGGWTRHATGELVGTGGRVAVPDWSAEGAEIALDEFYVALAGNFEYGP 2869
QY 66 ----- 65
Db 2870 LFOGQAAMRRGDEVLAETAPAEADAMASGYLLDPALLDAALQASALCDRPEQGGWLP 2929
QY 66 ----- 65
Db 2930 FSFTGVLSAPACTISRVLETRRPDAISVAVMDESGRLLASIDSRLRSVSSGQLANRD 2989
QY 66 ----- 65
Db 2990 AVRDALFEVTWEPVATQSTEPGRWALLGDTACGKODLIKLATDSADRCADLAALAEKLD 3049
QY 66 ----- 65
Db 3050 SALVPDVVVCAGEADPGTGAALAEQTQTLALLQAWLAEPRLAEARLVVVTCAAVTTA 3109
QY 66 ----- 65

Db 3110 PSDGASELAHALWGLLRAAOVENPQOFVLADVDGTAESWRALPSALGSMEPQALRKCA 3169
QY 66 ----- 65
Db 3170 VRAPRLASVAGOIDVPAVVADPDRTVLISGGTGLLGAVARHLVTERGVRRLLVLTGRROW 3229
QY 66 -APNIT-----ANLTSSLSLV-----CG----- 82
Db 3230 DAPGITELVGLNGLVAVVDVACVADRADLESLLAAVPABFFPLCGVVVHAAGALADGVI 3289
QY 83 -----WS----- 84
Db 3290 ESLSPDDVGAVFKAAGAWNLHELTRDTLSFPALFSSLSGVACAPGQGVYAAANAFLD 3349
QY 85 ----- 84
Db 3350 ALAHYRRSQGLPAVSLANGLWEQPSGMTETLSEVDRSRIARANPPLSTKEGLRFLDAGLA 3409
QY 85 ----- 84
Db 3410 LDRAAVVPKLDRTFLAEQARSGSIPALLTALVPIRRNRASCTELADEGTLGCVVREH 3469
QY 85 ----- 84
Db 3470 AAALGYSSAADVGVERAFRDLGFDLSGLSVELNRLAGVLGVLPATAVFDYPTPRALAR 3529
QY 85 -----OTINPED----- 91
Db 3530 FLHQELADEIATTPAPVTTTTRAPVAEDDLVAIVGMGRCPFGQVSSPBELMRLVAGGVDAV 3589
QY 92 ----- 91
Db 3590 ADFPADRGWDIAGLFPDPPERAGTKYVREGAFLTDADRFDAGFFGISPREALAMDPOURL 3649
QY 92 ----- 91
Db 3650 LLELSWEAIERAGIDPQSLRGSRGTGVFAGLVMYHDYGARFASRAPEGFEGLNGSAGSVA 3709
QY 92 ----- 91
Db 3710 SGRIAYSGFEGPAVTVDTACSSSLVALHLAGOSLRSGECDLALAGGVTVMTPTGTFVEF 3769
QY 92 ----- 91
Db 3770 SRQGLAPDGRCKSFAESADGTGMEGAGLVLLERLSDARRNGHRVLAUVRGSAVNODGA 3829
QY 92 -----DTD 94
Db 3830 SNGLTAPNGPSQORVIOQALASAGLSVSDVDAVENAGTGTGLGDPTEAQAIAITYGRDRD 3889
QY 95 PG-----HAD----- 99
Db 3890 PGRPLWLGSVKSNIHTQAAAAGVVIKVMAMMRHGQLPRTLHVDAPSSQVDSWAGRVOL 3949
QY 100 ----- 99
Db 3950 LTENTPMPDSGRPCRVGVSSFGISGTNAHVILEQSTGQMDQAAEPDSSPVLDPVVPVWVV 4009
QY 100 -----LVLYITR----- 106
Db 4010 SGTPEALSQAATLATYLDQNVDSPLVDVSGISLAVTRSNALDERAVVLGSDRDTLLSGLN 4069
QY 107 -----FD----- 108
Db 4070 ALAAGHEAAGVVTGPVIGIGRTGTFVAGOGQWLMGRRLYSEFPAPAGAFDEACAELOA 4129
QY 109 ----- 108
Db 4130 NLGREGVVRDVFGSDLSLDRTLWAQSGLFALQVGLWELLGTWGVVRPSVLGHSVGELA 4189
QY 109 ----- 108

Db 4190 AAFAGVLSMAEAAELVAGRARLMOALPSGAMLAVSATSEARVGLLDGVDRVGAAYN 4249
Qy 109 ----- 108
Db 4250 APGSVVLSDRDVLDGIAGRDLGGQIRSRWLRSVSHAFSHRMDPMLAEFAELARSVDYRS 4309
Qy 109 ----- 108
Db 4310 PRLPIVTLTGNLDDVGMATPEYVWVQVRFVRFADGVQALVDQGVDTTIVELGPDGALS 4369
Qy 109 ----- 108
Db 4370 SLVQECVAESGRATGIPLVRRDRDEVRTVLDAQTHRGAGVWGSEFFAGTRATQVDLP 4429
Qy 109 ----- 108
Db 4430 TYAFQRQRYWLEPSDSDVGTGVLGAEPHLLGAVVPVAGGDEVLLTGRLSVGHPLAE 4489
Qy 109 ----- 111
Db 4490 HRLGEVVVPGTALLEMAWRAGSQVGCERVEELTLEAPLVLPERGAAVQLVAGPDEAG 4549
Qy 112 -----RG 120
Db 4550 RRSQLYSRGADEGDWRRTASGLLAQANAVPPADSTAMPDPGAGQVDLAERFVLAERG 4609
Qy 121 VT----- 122
Db 4610 LTYGVPVQGLRAAHRHGGDDIFAEAGSPDASGFGIHPALLDAAHMAWGASPDSEARLP 4669
Qy 123 -----QLG--- 125
Db 4670 FSWRGAQLYRAEGAALVRVLSPLSGAVSLTLVDATGRRVAAVESLSTRPVSTDQIGAGR 4729
Qy 126 -----GACSPW----- 132
Db 4730 GDQERLLHVEWVRSASAGMSLTSCAVGLGEPWHAALKTTGVQVESHADLASLATEVA 4789
Qy 133 ----- 132
Db 4790 KRGSAPGAVIVPCPRPMQELPTAARRATQQAAMQLQWLADDFVSTRILLTHRAVS 4849
Qy 133 ----- 132
Db 4850 AVAGEDVLDLVHAPLWGLVRSQAEBHPRDFALIDMDDERASOTALAEALTAGEAQLAVRS 4909
Qy 133 -----SCLITEDTG----- 141
Db 4910 GYVLAPRLGVKVSQGEAFRNDGTVLVTGGTGGGLGALLARHLVSAHGVRHLLASRRGL 4969
Qy 142 ----- 141
Db 4970 AAPGADELVALEQAGADVAVVCDSADRDLSLARLVASVPAENPLRVVWHAAGVLDGVL 5029
Qy 142 ----- 141
Db 5030 MSMSPERLDVLRPKVDAANYLHETRELGLSAFVLFSSVAGLFGGAGQSYYAAGNAFLD 5089
Qy 142 ----- 141
Db 5090 ALAHCRQAQGLPALSASGLWASIDGMAGDLAAADVERLSRAGIGPLSAPGGLAFDAAV 5149
Qy 142 ----- 141
Db 5150 GSDEPLAPVLDVEALRVQARSVOTRIPEMLHGMAMGSPRRPTFTSRVPEPLHERLAGLS 5209
Qy 142 ----- 141
Db 5210 EGERQVQVLRVADIAVVLGHGRSSVDVIEKPLAELGFDLSLTAEIERNRLATATGRLRP 5269
Qy 142 ----- 141
Db 5270 ATLAFDHGTAALAHQVCAQLGTATAPAPRRDTDDNDATEPVRSLFQQAAGRIIDGMDL 5329

Qy 142 ----- 141
Db 5330 VKVAAQLRPVFGSPGELESLEPKPVQLSRGPBELALVCMPALIGMPAAQOYARIAAGPRDV 5389
Qy 142 ----- 141
Db 5390 RDVSVIPMPGFIAGEPLPSAIEAVARTQAEAVLQEFAGGSFVLVGHSSGGWLAHEVAGEL 5449
Qy 142 -----FDLQVT 147
Db 5450 ERRGVVPAGVVLDTYIPGEITPRFSVAMAHRTYEKLATFTMDQDVGIT 5498

RESULT 6
US-09-335-409-5
; Sequence 5, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goslisch, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-5

Query Match 27.0%; Score 209; DB 3; Length 7257;
Best Local Similarity 1.5%; Pred. No. 3.1;
Matches 93; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;

Qy 1 AAG----- 4
Db 877 AGGVSWSGVPFTAGRRVPLPTYPQQRQRYWIEAPAGLGATADAALQAQFYRVDPWPM 936
Qy 5 -----ILH----- 7
Db 937 PRSSVDSRRARSQGWVLADRGVGCEAAAAALSSQGCSCAVLHAPAEASAVAEQVQALG 996
Qy 8 ----- 7
Db 997 GRNDQGVLYLWGLDVAVVEAGASAEVAKVTHLAAAPVLALIOALGTGPRSPRLWIVTRG 1056
Qy 8 ----- 7
Db 1057 ACTVGGEPAAPCOAALMGMRVAALEHPGSGWGLVDLPDEESPEVEALVAELSSDAE 1116
Qy 8 ----- 7
Db 1117 DOLAFROGRRDAARLVAAPPEGNAAPVLSAEGSYLVGTGGLGALLVRLVWVERGAGHL 1176
Qy 8 -----LE-----LLVAVG 15
Db 1177 VLISRHGLPDEEHWGRQDPPEVRARIAIEALEAQGARVTVAAVDVADAEAGMALLAAVE 1236
Qy 16 P-----DVFOAHQ----- 23
Db 1237 PPLRGVVHAAGLLDDGLLAHODAGRLARLVLRPKVEGAWLHTLTREOPLDLVLFSSASG 1296
Qy 24 ----- 23
Db 1297 VFGSIGGSYAAGNAFLDALADLRRTOGLAALSIAWGLMAEGMGSGAORREHEASGWA 1356

QY 24 ----- 23
Db 1357 MPTSRALAMEWLLGTRATORVVIQMDWAHAGNAPRDASGRFWRDLVTATKEASSAVP 1416
QY 24 ----- 23
Db 1417 AVERWRNASVETRYSALYBELVRGVAVGMFTDQGLDVRGFAEQGLDSLMAVEIRKRL 1476
QY 24 ----- 23
Db 1477 QGELGMPLSATLAFDPTVERLVEYLLSQALELQDRTDVRSLPATEDPIAIVGAACRF 1536
QY 24 ----- 29
Db 1537 PGVEDLESYWOLLTEGVVSTVPADRWNGADGRVPGSGEQRQTYVPRGGFLREVETP 1596
QY 30 ----- 29
Db 1597 DAAFFHISPREAMSLDPOORLLLEVSWEAIERAGODPSALRESPTGVFVGAGPNEYAERV 1656
QY 30 ----- 29
Db 1657 QELADEAAGLYSGTGNMLSVAGRLSFFLGLHGPTLAVDTACSSSLVALHLGCOSLRGE 1716
QY 30 ----- 29
Db 1717 CDQALVGVGNMLLSPKTFALLSRMHALSPGCRCKTFSADADGYARAEGCAVVLKRLSDA 1776
QY 30 ----- 29
Db 1777 QDRDPILAVIRGTAINHDGSSGLTVPSPQAEBALLRQALAHAGVVPADVDFVECHGTG 1836
QY 30 ----- 29
Db 1837 TALGDPPIEVRLSDVYGOARPADRPLILGAANKLGHWEPAAGLAGLLKAVLALGOEIP 1896
QY 30 ----- 29
Db 1897 AOPELGELNPLLPWEALPVAVARAAVWPRTDRPRFAGVSFGMSGTNAHVLEEAPAVE 1956
QY 30 ----- 29
Db 1957 LWPAAPERSAELLVLSKSECALDAQAARLEHLDHPELGLGDVAFSLATTRSAMNHL 2016
QY 30 ----- 29
Db 2017 AVAVTSREGLLAALSVAOQOTPGAARCTIASSRGLAFLFTGQGAQTPGMGRGLCAAW 2076
QY 30 ----- 29
Db 2077 PAPREAFDRCVALFDRELDRLREVWMAEPGSAESLLLDQTAFTQPALFTVEYALTALWR 2136
QY 30 ----- 29
Db 2137 SWGVEPELVAGHSAGELVAACVAGVFSLEDGVLVAAGRLMQGLSAGAMVSLGAPAE 2196
QY 30 ----- 29
Db 2197 VAAAVAPHAASVIAAVNGPEQVVIAGVEQAVQAIAGFAARGARTKRLHVSHASHPLM 2256
QY 30 ----- 35
Db 2257 EPMLEBFGRVAASVTYRRPSVLSVNLGKVVADLSAPGYWVRHVREAVRFADGVKALH 2316
QY 36 ----- 35
Db 2317 EAGAGTFVEGPKPTLLGLLPACLPPEPTLLASLRAGREEAAGVLEALGRLWAAGGSVS 2376
QY 36 ----- 35
Db 2377 WPGVFTAGRRVPLPTYWQORQYWPDIPEPSRRHAAADPTQGFYRVDPWEIPRSLQKS 2436
QY 36 ----- 35

Db 2437 EASRGSWLVADKGVGEAVAAALSTRGLPCVVLHAPAEATSATAELVTEAAGRSWQV 2496
QY 36 ----- 39
Db 2497 VLYLWGLDVAVVGAEASIDEIGDATRRATAPVLGLARFLSTVSCSPRLVYVTRGACTVGDE 2556
QY 40 ----- 39
Db 2557 PAIAPQAAALWGMGRVAALHPGAMGGLVDLPDPRASPPPOASPIDGEMLVTELLSOETEDO 2616
QY 40 ----- 39
Db 2617 LAFRHRRHAARLVAAPPOGAAPVLSAEASYLVGTGLGGLGLIVAQWLVELGARHLVL 2676
QY 40 ----- 39
Db 2677 TSRRGLPDRQANCEOQPPEIRARIAAVEALEARGARVTVAAVDVADVEPMTALVSSVBP 2736
QY 40 ----- 39
Db 2737 LRGVVHAAGVSYMRPLAETDETLLSVLRPKVAGSWLLHRLHGRPLDLFVLFSSGAAMV 2796
QY 40 ----- 39
Db 2797 GSHSOGAYAAANAFLDGLAHLRRSOSLPALSVAMGLMAEGOMADAHAHRLSDIGVLPMS 2856
QY 40 ----- 39
Db 2857 TSAALSALQRLVETGAAQRTVTRMDWARFAPVYTAGRRNLLSALVAGRDIAPSPPAA 2916
QY 40 ----- 39
Db 2917 TRNWRGLSVABARVALHEIVHGAVARVLGFLDPSALDPGMFNGEQLDSLMAVEIRNLQ 2976
QY 40 ----- 39
Db 2977 AELDVLSTTLAFDHPPTVQRLVEHLVDVLKLEDRSOTÖHVRSLASDEPIAIVGAACRFP 3036
QY 40 ----- 39
Db 3037 CGVEDLESYWOLLAEVGVSAEVPADRDAADWYDPDPEIPGRYVTKGAFRLDLQLDA 3096
QY 40 ----- 39
Db 3097 TFFRISPREAMSLDPOORLLLEVSWEALESAGIAPDLRDSPTGTVFVGAGPNEYVYTORLR 3156
QY 40 ----- 39
Db 3157 GFTDGAAGLYGGTGNMLSVTAGRLSPFLGLHGPTLAMDTCSSSLVALHACOSLRLGEC 3216
QY 40 ----- 39
Db 3217 DQALVGVNVLLAPETFTVLLSRMRALS PDGRCKTFSADADGARGEGCAVVLKRLDAQ 3276
QY 40 ----- 42
Db 3277 RAGDSILALIRGSANVHDGPSSGLTVPNGPAQOALLRQALSQAGVSPVDVDFVECHGTGT 3336
QY 43 ---DP--- 44
Db 3337 ALGDPTEVOALSEVYGPGRSGDRPLVLGAANKANVAHLEAAAGSLALLKAVLALRHEQIPA 3396
QY 45 ----- 44
Db 3397 QPELGELNPHLPWNTLPVAVPRKAVPWGRGRPRRAGVSAGLSGTNVHVHVEAEPEVP 3456
QY 45 ----- 44
Db 3457 APAAPRPVELVVLSAKSAALDAAAARLSAHLSAHPELSGLDVAFSLATTRSPMEHRLA 3516
QY 45 ----- 44

Db 3517 IATTREALGALDAAAQOKTPOGAVRGKAVSSRGKLAFLFTCGAQMCGMGRGLYETWP 3576
Qy 45 ----- 44
Db 3577 AFREAFDRCVLFDREIDQPLREVMMAPGLAQAAARLDQYAPALFALEAALMRS 3636
Qy 45 ----- 44
Db 3637 WGVEPHVLLGHSIGELVAACVAGVFSLEDAVLVAARGRLMQALPAGGAWVAIASEAEV 3696
Qy 45 -----SLGAQF----- 50
Db 3697 AASVAPHAATVSIAAVNGPDVAVIAGAEVQVIALGATFAARGIRTKLAVSHAFHSPLMD 3756
Qy 51 ----- 50
Db 3757 PMLEDQFQVNAATYRAPDRPVVSNVTGHVAGPEIATPEYVVRHRSVAVRFGDGAKALHA 3816
Qy 51 ----- 50
Db 3817 AGAATFEVGPKEPVLGLLPACLGADAVLVPSLRADRSECEVVLAAALGAWYANGGALDW 3876
Qy 51 -----RV----- 52
Db 3877 KGVFPDARRVALPMYPWORERHMDLTPRSAAAPAGIAGRWPLAGVGLCMGAVLHVLS 3936
Qy 53 -----HLV----- 55
Db 3937 IGRPHOPFLGDHLVFGKVVVPGAFHVAVILSTAERWPERAIELTGVBEFLKAIAPEPOE 3996
Qy 56 ----- 55
Db 3997 VELHAVLTPEAAGDGYLFELATLAAPETERMTTHARGRVQTDGAPCALRLEVLDBA 4056
Qy 56 ----- 55
Db 4057 IQPLDFAGFLDRLSAVRIGWGPLWRWLQDGRVGDSEASLTLVPTYPNAHDVAPLPIILD 4116
Qy 56 ----- 55
Db 4117 NGFAVLLSTRSEPDGTPPLPFAVERVWRWRAVPGRVRCGGVPRSOAFGVSFVLVDE 4176
Qy 56 ----- 55
Db 4177 TGEVVAEVEGFVCRRAPREVFLQESGASTAALYRLDWPEAPLPDAPAEIEESWVVAA 4236
Qy 56 ----- 55
Db 4237 PGSENAALATRLNRCVLAEPKGLEAALAGVSPAGVICLWEAGAHEEAPAAAQVATEGL 4296
Qy 56 ----- 55
Db 4297 SVVQALDRVRLMWVTMGAVAEAGERVQVATAPVNGLGRVTMQERPELSCTILVDLEPE 4356
Qy 56 -----KWVLTPEG----- 65
Db 4357 ADAARSADVLLRELGRADDETOVAFRSGKRVARLVKATTPGELLVPDAESYRLEAQKG 4416
Qy 66 ----- 65
Db 4417 TLDQLRLAPQRRAPGPEVEIKVTASGLNFTVLAVLGMYPGDAGPMGGDCAGVATAVG 4476
Qy 66 ----- 65
Db 4477 QGVRHVAVGDVMTLGTLHREFTVDARLVVRQAPAGLTPAQAAATVPVAFLTAWLALHDLGN 4536
Qy 66 ----- 65
Db 4537 LRRGERVLIHAAAGGVMAAVQIARWIGAEVEFATASPKWAQVQAMGVPRTHIASRTLE 4596
Qy 66 ----- 65
Db 4597 FAETFRQVTGGRGVDVVLNALAGEFVDASLSLLSTGGRFLEMCKTDIRDBAAVAAAHPGV 4656

Qy 66 ----- 65
Db 4657 RYRVFDILELAPDRTRILERVVEGFAAGHLRALPVHFAITKAEAAFRFMAQARHOKV 4716
Qy 66 ----- 65
Db 4717 VLLPAPSAAPLAPTTGTLTGLGALGLHVARMLAQOVPHMVLTGRRGLDTPGAAKAVA 4776
Qy 66 ----- 65
Db 4777 EIEALGARVTIAASDVADRNALEAVLOAIPEMPLOQVTHAAGALDGVLDDEOTDRFSR 4836
Qy 66 --APNIT----- 70
Db 4837 VLAPKVTGAWNLHELHTAGNDLAFVFLPSSMSGLGSAGQSNYAAANTFLDALAAHRAEG 4896
Qy 71 -----ANLTSS----- 77
Db 4897 LAAQSLANGPWSOGMAAGLSAALQARLARHGMGALSPOGTALLGOALARPETOLGAMS 4956
Qy 78 ----- 77
Db 4957 LDYRAASQASGAAPVPMWRALVRABARHTAAGAGOGAALARLALGALPEARRADEVKRYQAE 5016
Qy 78 ----- 77
Db 5017 IARVLSMAASAVPDRPLSLDGLSLTAVELRNVLQORVGATLPATLAFDHTVDALTR 5076
Qy 78 ----- 77
Db 5077 WLLDKVLAEPVSSAKSSPOVALDEPIAIIIGICRPPGVADPEFWRLLEEGSDAYV 5136
Qy 78 ----- 77
Db 5137 EVPHERWDIDAFYDPDPDVRGKMTTRFGGFLSDIDRFDPAPFGISPREATTMDPOORLLL 5196
Qy 78 ----- 77
Db 5197 ETSWEAFERAGILPERLMGSDTGTVFGLFYQBYAALAGGIEAPDGYLGTGTASVASGRI 5256
Qy 78 ----- 77
Db 5257 SYVLGLKPSLTDITACSSSLVAVHLAQALRRGECVSVALAGGVALMLTPTATFVEFSRLR 5316
Qy 78 ----- 77
Db 5317 GLAPDGRCKSFSAADGVGWSEGCAMLLKPLRDAQRDGPILAVIRGTAVNQDGRSNGL 5376
Qy 78 ----- 77
Db 5377 TAPNGSSQOEVIIRRAEOAGLAPADVSYVECHGTGTTLLGDPLEVOALGAVLAQGRPSDRP 5436
Qy 78 ----- 77
Db 5437 LVIGSVKSNIGHTQAAAGVAGVIKVALALERGLIPRSLHFDAPNPHIPWSELAVOVAAKP 5496
Qy 78 ----- 77
Db 5497 VEWTRNGVPRRAGVSSFGVSGTNAHVLEEAPAAAFAPAAARSSELVLSAKSAAALDAQ 5556
Qy 78 ----- 77
Db 5557 AARLSAHVVAHPPELGLGLDGLAFSLATTRS PMTYRLVAVAATSREALSAAALDTAAOQAPPA 5616
Qy 78 -----LSVC-----GWS----- 84
Db 5617 ARGHASTGSPKVVVFPFGQSSQMLGMQKLLSEEPFRDALSACDRAIQAEAGWSLLAE 5676
Qy 85 ----- 84
Db 5677 LADETTSQLGRIDVQFALFAIEVALSALMRSWGPEDAVVGHSMGEVAAAHVAGALS 5736

```
QY 85 ----- 84
Db 5737 EDVAIIICRRSLLRRISQOGEMAVVELSLAEAEALLGYEDRLSVAVSNSPRSTVLAGE 5796
QY 85 ----- 84
Db 5797 PAALAEVLAILAAKGVCFCRVRKVDVASHSPQIDPLRDELLAALGELEPRQATVSMRSTVT 5856
QY 85 ----- 84
Db 5857 STIMAGPELVASYADNVQRQVFAEAVQSLMEDGHGLFVEMSPHILTTTSVEIRATK 5916
QY 85 ----- 84
Db 5917 REGVAVGSLRRGQDERLSMLBALGALVWHGQAVQWNERLFSAGGAGLRVRPLPTVWQROR 5976
QY 85 ----- 90
Db 5977 YWVDAPTGAAGGSFHAHAGSHPLLGEMOTLSTQSTRVWETTLDLKRLPWLGDHRVQGA 6036
QY 91 ----- 94
Db 6037 VVFGAAVLEMAISGAELGDPLOVSDVVLAELAFADDTFAAVQVMATEERPGRLOF 6096
QY 95 ----- 97
Db 6097 HVASRVFGHGAAPRSHARGVLRQIERAEVPAERLDLAALRLOASAPAAATYAALAEWG 6156
QY 98 ----- 97
Db 6157 LEYGPAPQGLVELWRGEALGRVRLPEAAGSPAACRLHPALLDACPHVSSAFADRGAT 6216
QY 98 ----- 97
Db 6217 PWDPVEIGSLRWFORPSEGLWCHARSVSHGKPTDRRSTDFWVVDSTGAIVAEISGLVAQ 6276
QY 98 ----- 97
Db 6277 RLAGVRRREDDWFMEPAWPTAVPGSEVMAGRWLLIGSGGGLGAALHSALTEAGHSVV 6336
QY 98 ----- AD 99
Db 6337 HATGRGTSAAAGLQALLTASFDGOAPTSSVHLGSLDERGVLDADAPDADALEESLVRGCD 6396
QY 100 LVLV-----ITR----- 106
Db 6397 SVLWTVQAVAGAFRDPRLWLVTGCAIGAGDVSVAQAPLLGLGRVIALEHAELRCAR 6456
QY 107 ----- 106
Db 6457 IDLDPARRDGEVDELLAELLADDAEEVAFRGGERRVARLVRLPETDCREKIEPAEGRP 6516
QY 107 FDLEL----- 111
Db 6517 FRLEIDSGVLDLVLATERPPGPEVEIAVEAAGLNFLDVMRAMGIYPGDPGPVAL 6576
QY 112 ----- 111
Db 6577 GAECGRIVAMGEGVESLRIGQDVAVAFPSFGTHVTIDARMLAPRPAALTAQAALPVP 6636
QY 112 ----- 111
Db 6637 AFMTAWYGLVHLGRLAGERVLIHSATGGLAAVQIARHGLAEIFATAGTPEKRAWLRE 6696
QY 112 ----- PDGN----- 115
Db 6697 QGTAHVMDRSRSLDFAEOVLAAATKGECDVVVLSLSGAATDASLSLTPDGRFIELGKTDI 6756
QY 116 ----- 115
Db 6757 YADRSGLAHFRKLSYSADVDLAGLAVRRPVERVAALLAEVVDLLARGALQPLPVEIFFLS 6816
QY 116 ----- 115
```

```
Db 6817 RAADAFRMAQAOHLGKLVLALEDPDVRIRVPGESGVAIRADGAVLVTGGGLGLSVAG 6876
QY 116 -----ROV----- 118
Db 6877 WLAEOGAGHLVLVGRSGAVSAEQQTAAVALEAHGARVTVARADVADRAQMERILREVTA 6936
QY 119 -----RGVTQGG----- 126
Db 6937 GMPLRGVVHAAGILDGGLMQOTPARFRAVMAPKVRGALHLHALTREAPLSFFVLVASGA 6996
QY 127 -----ACSTWSQL----- 135
Db 6997 GLLGSPGOGNYAAANTFLDALAHRRRAOGLPALSIDWGLFADVGLAAGQONRGARLVTRG 7056
QY 136 -----ITEDTG-----FD----- 143
Db 7057 TRSLPDEGLWALERLLDGRDRTQAGVMPFDVRQWVEFFPAAASSRRLSRLMTARRVASGR 7116
QY 144 ----- 143
Db 7117 LAGDRDLLERLATAEAGARAGMLQEWRAQVSOVLRLSEGKLDVDAPLTLGLMDSLMGLE 7176
QY 144 -----LGVTI 148
Db 7177 LNNRIEAVLGITM 7189

RESULT 7
US-09-568-102-5
; Sequence 5, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-102-5

Query Match 27.0%; Score 209; DB 4; Length 7257;
Best Local Similarity 1.5%; Pred. No. 3.1;
Matches 93; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;

QY 1 AAG----- 4
Db 877 AAGGSVMPGVFTAGRRVPLTPYPMQRQRYWIEAPAEGLGATAADALAQMRYRVDWPEM 936
QY 5 -----ILH----- 7
Db 937 PRSSVDSRRARSRGWLVLADRGVGEAAAAALSSQSCSCAVLHAPAEASAVAEQVTOALG 996
QY 8 ----- 7
Db 997 GRNDWQGVLYLWGLDAVVEAGASAEVAKVTHLAAAPVLALIOALGTGPRSPRLWIVTRG 1056
QY 8 ----- 7
Db 1057 ACTVGGEPDAAPCOAALWGMGRVAALEHPGSGMGLVDLPDEESPTVEALVAELLSPDAE 1116
```

```
QY 8 ----- 7
Db 1117 DOLAFQGRRRRAARLVAAPPEGNAAPVLSAEGSYLVTTGGTGLGALGLVARWLVERGAGHL 1176
QY 8 -----LB-----LLVAVG 15
Db 1177 VLISRHGLPDREWGDRDQPEVVRARIAIAEALQAQARTVVAADVADAEAGMAALLAAVE 1236
QY 16 P-----DVFOAHO----- 23
Db 1237 PPLRGVVHAAGLLDDGLLAHQDAGRLARVLPRKVEGAWVLHLLTREOPLDLFLVFSSASG 1296
QY 24 ----- 23
Db 1297 VFGSICGSAAGNAFLDALLRRTOGLAALSIAWLAEGMGSAQORREHEASGIWA 1356
QY 24 ----- 23
Db 1357 MPTSRALAAWELLGTRATORVVIONDWAHAGAAPRDSRGRFWRDLVTATKEASSAVP 1416
QY 24 ----- 23
Db 1417 AVERWRNASVETRYSALVELVRGVVAGVMGFTDQGLDVRGFAEQGLDSLMAVEIRKRL 1476
QY 24 ----- 23
Db 1477 OGELGMLSATLAFDHTVERLVEYLLSQALELQDRDVRSLRPATEDPIAIVGAACRF 1536
QY 24 -----EDTERY----- 29
Db 1537 PGVEDLESYWQLLTFEGVVVSTVPADRMNGADGRVPGSGEAQRQTYVPRGGFLREVETF 1596
QY 30 ----- 29
Db 1597 DAAFFHISPREAMSLDPQORLLLEVSWEAIERAGQDPSALRESPTGVFVGAGPNEYAERV 1656
QY 30 ----- 29
Db 1657 QELADEAAGLYSGTGNMLSVAAGRLSFFLGLHGPTLAVDTACSSSLVALHLCQSLRRGE 1716
QY 30 ----- 29
Db 1717 CDQALVGVNMULLSPKTFALLSRMHALSPGGRCKTFSADADGYARAECAVVVLKRLSDA 1776
QY 30 ----- 29
Db 1777 QRDRDPILAVIRGTAINHDGPSSGLTVPSGPAQEALLRQALAHAGVVPADVDFVECHGTG 1836
QY 30 ----- 29
Db 1837 TALGDPIEVRALSDVYGOARPADRPILGAANKANLGHMEPAAGLAGLLKAVLALCQEIQIP 1896
QY 30 ----- 29
Db 1897 AQPELGELNPLLPWEALPVAVARAAPVMPRTDRPRFAGVSSFGMSGTNAHVVLEEPAVE 1956
QY 30 ----- 29
Db 1957 LWPAAPERSAELLVLSKSEGDALDAQAARLREHLDMHPELGLGDVAFSLATRSMMNHLR 2016
QY 30 ----- 29
Db 2017 AVAVTSREGLLAALSAVAQGTTPGAARCIASSSRGKLAFLFTGQAGTQPGMGRGLCAAW 2076
QY 30 ----- 29
Db 2077 PAFREAFRCVALFDRELDRLPRLVWMAEPGSAESLLLDQTAFTQPALFTVEYALTALWR 2136
QY 30 ----- 29
Db 2137 SKGVEPELVAGHSAGELVAACVGVFSLEDGVLVVAARGRLMQGLSAGGAMVSLGAPEAE 2196
QY 30 ----- 29

QY 2197 VAAAVAPHAASVSIAAVNGPEQVVIAGVEQAVQAIAAGFAARGARTKELHVSHASHSPLM 2256
QY 30 -----VLTNLN----- 35
Db 2257 EPMLEEFGRVAASVTYRRPSVLSVNSLSGKVVVADELSAPGYWVRHVRREAVRFADGVKALH 2316
QY 36 -----:::|----- 35
Db 2317 EAGAGTFVEVGPKPTLLGLLPACLPEAEPTLLASLRAGREAAAGVLEALGRLWAAGGSVS 2376
QY 36 ----- 35
Db 2377 WPGVPTAGRRVPLPTYWQRQRYWPDIEPDSRRRAAADPTQGWFYRVDMPEIPIRSLQKS 2436
QY 36 ----- 35
Db 2437 EASRGSLWLVLADKGGVGEAVAAALSTRGLPCVVLHAPETSATAELVTEAAGGRSDHQV 2496
QY 36 -----IGAE----- 39
Db 2497 VLYLWGLDAAVVGAEASIDEIGDATRRATAPVLGLARFLSTVSCSPRLWVVTGACIVGDE 2556
QY 40 ----- 39
Db 2557 PAIAPCQAAALWGMGRVAALHPGAWGGLVDLDPRASPPQASPIDGEMLVTELLSQETEDQ 2616
QY 40 ----- 39
Db 2617 LAFRHRRAARLVAAPPQGOAAPVLSAEASYLVTTGGLGGLIIVAQWLVELGARHLVL 2676
QY 40 ----- 39
Db 2677 TSRRGLPDRQWCEQQPPEIRARIAAVEALEARGARVTVAADVADVEPMTALVSSVEPP 2736
QY 40 ----- 39
Db 2737 LRGVVHAAGVSMRPLAETDETLLSVLRPKVAGSMLLHRLHGRPLDLFLVFSSGAAVW 2796
QY 40 ----- 39
Db 2797 GSHSQGAYAAANAFDGLAHLRRSQSLPALSVANGLWAEGMADAHAHARLSDIGVLPM 2856
QY 40 ----- 39
Db 2857 TSAALSALQRLVETGAARTVTRMDWARFAPVYTARGRRNLLSALVAGRDI IAPSPAAA 2916
QY 40 ----- 39
Db 2917 TRNWRGLSVAEARVALHEIVHGAVARVLGFLDPSALDPGMPGNEOGLDSLMAVEIRNLLQ 2976
QY 40 ----- 39
Db 2977 AELDVLSTTLAFDHPVTQRLVEHLLVDVLKLEDRSDTOHRVSLASDEPIAIVGAACRFP 3036
QY 40 ----- 39
Db 3037 GGVEDLESYWQLLAEGVWVSAEVPADRWDAADWYDPPPEIPGRTYVTYTKGAFRLDQLRDA 3096
QY 40 ----- 39
Db 3097 TFFRISPREAMSLDPQORLLLEVSWEALESAGIAPDTRDSPTGVFVGAGPNEYVYTORLR 3156
QY 40 ----- 39
Db 3157 GFTDGAAGLYGGTGNMLSVTAGRLSPFLGLHGPTLAWDTACSSSLVALHACOSLRUGE 3216
QY 40 ----- 39
Db 3217 DQALVGVNVVLLAPEITFVLLSRMRALS PDGRCKTFSADADGYARGEGCAVVVLKRLDAQ 3276
QY 40 -----LUR----- 42
|||
```

```
Db 3277 RAGSILALIRGSVNHDBGSSGLTVPGPAQOALLRQALSQAGVSPVDVFVECHGTGT 3336
Qy 43 ---DP-----
||
Db 3337 ALGDPIEQALSEVYGPGRSDRPLVLGAAKANVAHLEAAAGSLALKAVLALRHEQIPA 3396
Qy 45 -----44
Db 3397 QPELGELNPHLPWNLTLPVAVPRKAVPMGRGARPRRAGVSAGFSGTNNVHVLEAEPEV 3456
Qy 45 -----44
Db 3457 APAARPVELVLSAKSAAALDAAARLSAHLSAHPELSIGDVAFLATTRSPMEHRLA 3516
Qy 45 -----44
Db 3517 IATTSREALGALDAAAQKTPOGAVRGKAVSRGKLAFLTGQAGMPGMRGLYETWP 3576
Qy 45 -----44
Db 3577 AFREAFDRCVALFDREIDQPLREVMWAAPOGLAQARLDQTYAQPALFALEYALALWRS 3636
Qy 45 -----44
Db 3637 WGEVPHVLGHSIGELVAACVAGVFSLEDAVRLVAARGRLMQALPAGGAMVAIAASEAEV 3696
Qy 45 -----50
|||
Db 3697 AASVAPHAATVSIAAVNGPDVAVIAGAEVQVLALGATFAARGIRTKRLAVSHAFHPLMD 3756
Qy 51 -----50
|||
Db 3757 PMLEDFORVAATTAYRAPDRPVVSNVTGHVAGPEIATPEYVWRHRSVRPFGCAKALHA 3816
Qy 51 -----50
Db 3817 AGAATFVEVGPKPVLGLLPACLGEADAVLPSLRADRSECEVLAALGAWAYAWGALDW 3876
Qy 51 -----52
||
Db 3877 KGVPDGGARRVALPMYQWERHMDLTPRSAAPAGIAGRWPAGVGLCPGAVLHHVLS 3936
Qy 53 -----55
|||
Db 3937 IGRHQPFGLDHLVFGKVVVPGAFHVAVILSIAAERPRAIELTGVFELKATAMPDQE 3996
Qy 56 -----55
Db 3997 VELHAVLTPEAAGDGYLFELATLAAPEPERRTHTHARGVQPTDGPAGALPRLEVLDDRA 4056
Qy 56 -----55
Db 4057 IQPLDFAGFLDRLSAVRIGWGLWRMLQDQGVGEASLATLVPTYPNABHDVAPLHPILLD 4116
Qy 56 -----55
Db 4117 NGFAVLSLSTRSEDDGTPLPFAVERVWRAPVGRVRCGGVPRSQAFGVSSFVLVDE 4176
Qy 56 -----55
Db 4177 TGEVVAEVEGVCRRAPREVFLRQESGASTAALYRLDWPEAPLPDAPABRIEBSWVWAA 4236
Qy 56 -----55
Db 4237 PGSEMAAALATRLNRCVLAEKPLEAALAGVSPAGVICLWEAGAHEEAPAAQORVATEGL 4296
Qy 56 -----55
Db 4297 SVVQALDRRAVRLWVTMGAVAVEAGERVQVATAPVNWGLGRVTWQERPELSCTLVLDLEF 4356
Qy 56 -----65
|||
Db 4357 ADAARSADVLLRELGRADDETQVAFRSGKRRVARLVKATTPTEGLLVDAESYRLEAGQKG 4416
|||
Qy 66 -----65
Db 4417 TLDQLRLAPQARRAPGPEVEIKVTASGLNFRTVLAVLWMPGDAGPMGGDCAGVATAVG 4476
Qy 66 -----65
Db 4477 QQVRHVAVGDAVMTLGTTLHRTVTVDARLVVRQPAGLTTPAQAAATVPVAFLTAWLALHDLGN 4536
Qy 66 -----65
Db 4537 LRRGERVLIHAAAGVGMAAQIARWIGAEVPATASPSKWAQVQAMGVPRTHIASRRTLE 4596
Qy 66 -----65
Db 4597 FAETFRQVTGGRGVDVNLALAGEFVDASLSLLSTGGRFLEMKTDIRDRAVAAAHPGV 4656
Qy 66 -----65
Db 4657 RYRVFDILELADPDRTREILERVVEGFAAGHLRALPVHAFATKABAAFRFMAQARHOGKV 4716
Qy 66 -----65
Db 4717 VLLPAPSAAPLAPTGTVLLTGGLGALGLHVARWLAQOQVPHMVLTRGRGLDTPGAAKAVA 4776
Qy 66 -----65
Db 4777 EIEALGARVTIAASDVADRNALAVLOAIPAEMWPLQGVIIHAAGALDDGVLDQTTDRFSR 4836
Qy 66 ---APNIT-----70
|||
Db 4837 VLAPKVTGAWNLHELTAGNDLAFVLFSSMSGLLGSAGOSNYAANTTFLDALAAHRAEG 4896
Qy 71 -----ANLTSSL-----77
|||
Db 4897 LAAQSLAWGPWSGCGMAAGLSAALQARLARHGMGALSPOGTALLGOALARPETQLGAMS 4956
Qy 78 -----77
Db 4957 LDVRAASOASGAAPPVPMWRALVRABARHTAAGAOGALAARLALPEARRADEVRKVQAE 5016
Qy 78 -----77
Db 5017 TARVLSWASAASAVPVDRLPSDLGLDSLTAVELRNVLQORVGATLPATLAFDHTPTVDALTR 5076
Qy 78 -----77
Db 5077 WLLDKVLAVABSPVSSAKSSPOVALDEPIAIIIGICRPPGVAOPESFWRLLLEGSDAVV 5136
Qy 78 -----77
Db 5137 EVPHERWDIDAFYDPPDPVRGKMTTRFCGFLSDIDRFPAPFGISPREATTWDPQORLLL 5196
Qy 78 -----77
Db 5197 ETSWEAFERAGILPERLMGSDTVGVGLFYOEYAAALAGGIEAFDGYLGTGTGTASVASGR 5256
Qy 78 -----77
Db 5257 SYVGLKGPSLTVDTACSSSLVAVHLAQALRRGECVALAGGVALMLTPATFVFSRLR 5316
Qy 78 -----77
Db 5317 GLAPDGRCKSFSAADGVGWSEGCAMLLKPLRDAQRDGPILAVIRGTAVNOGRSNGL 5376
Qy 78 -----77
Db 5377 TAPNGSSQOEVI RRALEOAGLAPADVSYVECHGTGTTLLGDPLEVOALCAVLAOGRPSDRP 5436
Qy 78 -----77
Db 5437 LVIGSVKSNIGHTQAAAGVAGVIKVALALERGLIPRSLHFDAPNPHIPWSELAVOVAAPK 5496
```


QY	78	-----	77
Db	5497	VEWTRNGVRRAGVSSFGVSGTNAHVLEEAPAAAFAPAAARSAEFLVLSAKSAAALDAQ	5556
QY	78	-----	77
Db	5557	AARLSAHVVAHPGLGLGLDGLAFSLATTSRPMYRLAVALTSREALSALDAAQOQAPPA	5616
QY	78	-----	84
Db	5617	ARGHASTGSAKVVVFPQGSQWLGKQKLLSEEPFRDALSCADRAIOAEGWLLAE	5676
QY	85	-----	84
Db	5677	LADETTSQLGRIDVQPALFAIEVALSALWRSWGVEPDAVVGHSMEVAHAHVACALS	5736
QY	85	-----	84
Db	5737	EDAVAIICRRSLRRISGOGEMAVVLSAEAEALLGYEDRLSVAVSNSPRSTVLAGE	5796
QY	85	-----	84
Db	5797	PAALAEVLAIAAKGVFCRRVKVDVASHSPQIDPLRDELLAALGELEPRQATVSMRSTVT	5856
QY	85	-----	84
Db	5857	STIMAGELVASYNADNVQRPVFAEAVOSLMEDGHGLFVEMSPHILTTTSVEEIRATK	5916
QY	85	-----	84
Db	5917	REGVAVGSLRRGODERLSMLEALCALVWVGQAVGWERLFSAGAGLRVPLPTYPQORER	5976
QY	85	-----	90
Db	5977	YWVDAPTGAAGRSFEAHAGSHPLLGMOTLSTORSTRVWEITLDLKRPLWLGDRVQGA	6036
QY	91	-----	94
Db	6037	VVFPGAAYLEMALSGAEALGDGQLQVSDVLAELAFADDTTAAVQVMATEERPGRLOF	6096
QY	95	-----	97
Db	6097	HVASRVPGHGAARSHARGVLRQIERAEVPAERLDAALRALQASAPAAATYAALABMG	6156
QY	98	-----	97
Db	6157	LEYGPAFOGLVLRGEALGRVLRPEAAGSPAACRLHPALLDACFHVSSAFADRGAT	6216
QY	98	-----	97
Db	6217	PHVPEIGSLRWFQPSGELMCHARSVSHGKPTDPRRSTDPWVVDSTGAIVAEISGLVAQ	6276
QY	98	-----	97
Db	6277	RLAGVRRREDDWFMPEAPTEAVPGSEVMAGRWLLIGSGGLGAALHSALTGAHSV	6336
QY	98	-----	AD 99
Db	6337	HATGRGTSAGLQALLTASFQCAPTSVVHLGSLDERGVLDADAPFDADALESLVRGCD	6396
QY	100	LVLY-----	106
Db	6397	SVLWTVQAVAGAFRDPPLRLVTRGAQIGAGDVSVQAAPLLGLGRVIALEHAELRCAR	6456
QY	107	-----	106
Db	6457	IDLOPARRDGEVDELLAELLADDAEEVAFRGERRVARLVRLPETDCREKIEPAGRP	6516
QY	107	FDLEL-----	111
Db	6517	FRLEIDSGVLLDLVLRATERPPGCEVEIAVEAAGLNFLDVMRWANGIYPGGDPVAL	6576
QY	112	-----	111

Db	6577	GAECGRIVAMGEGVESLRIGQDVVAVAPPSFGTHVTIDARMLAPRPAALTAQAALPV	6636
QY	112	-----	111
Db	6637	AFMTAWYGLVHLGRIRAGERVLIHSATGCTGLAAVQIARHLGAEIFATAGTPEKRAWLRE	6696
QY	112	-----	115
Db	6697	OGIAHVMDSRSLDFAEQVLAATKGEVDVVLNSLGAASLSTLVPDGRFIELGKTDI	6756
QY	116	-----	115
Db	6757	YADRSGLAHFRKSLSYSVADLAGLAVRRPERRVAALLAEVVDLLARGALOPLPVEIFPLS	6816
QY	116	-----	115
Db	6817	RAADAFRMAQAQHLGKLVLALEDPDVRIIRVPGESGVAIRADGAYLVTTGGLGLSLSVAG	6876
QY	116	-----	118
Db	6877	WLAEQAGAGHLVLRSGAVSAEQQTAVAALEAHGARVTVARADVADRAQMERILREVTS	6936
QY	119	-----	126
Db	6937	GMPLRGVVHAAGILDDGLLMQOTPARFRAVMAPKVRGALHHLALTREAPLSFFVLVASGA	6996
QY	127	-----	135
Db	6997	GLLGSPQGNAYAAANTFLDALAHHRAOGLPALSIDMGLFADVLGAAGQONRGARLVTRG	7056
QY	136	-----	143
Db	7057	TRSLTDEGLWALERLLDGDRTQAGVMPDFVQWVEFYPAASRRRLSRLMTARRVASGR	7116
QY	144	-----	143
Db	7117	LAGDRLLERLATAEAGAGAGMLQEVVRAQVSOVLRLSEGKLDVDAPLTSGLMDSLMGL	7176
QY	144	-----	148
Db	7177	LRNRIEAVLIGTM	7189

RESULT 8

US-09-567-969-5

; Sequence 5, Application US/09567969

; Patent No. 6355457

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goeblach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/567,969

; PRIOR APPLICATION NUMBER: 2000-05-10

; PRIOR FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 7257

; TYPE: PRT

; ORGANISM: Sorangium cellulosum

US-09-567-969-5

Query Match 27.0%; Score 209; DB 4; Length 7257;

Best Local Similarity 1.5%; Pred. No. 3.1;

Matches 99; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;

QY	1	AAGG-----4	Db	1957	LWPAAPERSABELLVLSGKSEGALDAQARLEHLDMPHLPGLGDVAFSLATTRSAMNHL	2016
Db	877	AAGGSVNPVGFPTAGRRVPLPTYPWQRQYWIETAPAEGLGATAADALAQWIFYRVDMPEM	QY	30	-----29	
QY	5	-----ILH-----7	Db	2017	AVAVTSREGLLAALSAVAQOTPPGAARCIASSSRGKLAFLLFTGGQATPGMGRGLCAAW	2076
Db	937	PRSSVDSRRARSOGWLVLADRGVGEAAALSSQGCSCAVLHAPAEASAVAEQVTOALG	QY	30	-----29	
QY	8	-----7	Db	2077	PAFREAFDRCVALFDRELDRLPREVMMAEPGSAESLLLDOTAFQTQPALFTVEYALTALWR	2136
Db	997	GRNDQGVLYLWGLDAYVEAGASAEVAKVTHLAAAPVLALIOALGTGPRSPRLWIVTRG	QY	30	-----29	
QY	8	-----7	Db	2137	SWGVEPELVAGHSAGELVAACVAGVFSLEDCGRLVLAARGRLMOGLSAGGAMVSLGAPAE	2196
Db	1057	ACTVGGEPDAAPCOAALWGMGRVAALEHPGSGWGLVDLDPESPTVEALVAELLSFDAE	QY	30	-----29	
QY	8	-----7	Db	2197	VAAAVAPHAASVSIAAVNGPEQVVIAGVEQAVQAIAGFAARGARTKRLHVSHASHSPLM	2256
Db	1117	DOLAFROGRRRAARLVAAPPEGNAAPVLSAEGSYLVTTGGLGALGLLVARWLVERCAGHL	QY	30	-----35	
QY	8	-----LE-----LLVAVG 15	Db	2257	EPMLEEFGORVAASVTYRRPSVSLVNSLSGKVADELSPGYVVRHVREAVRFADGVKALH	2316
Db	1177	VLISRHGLPREEWRDQPEVRIARIAIALEAQGARVTVAADVADAEAGMAALLAAVE	QY	36	-----35	
QY	16	P-----DVFOAHO-----23	Db	2317	EAGAGTFVEVGPKPTLLGLLPACLPEAEPTLLASLRAGREAAAGVLEALGRLWAGGSVS	2376
Db	1237	PPLRGVVHAAGLLDDGLLAHQDAGRLARLVRPKVEGANVWLHTLTREOPLDLFLVLFSSAG	QY	36	-----35	
QY	24	-----23	Db	2377	WPGVFPTAGRRVPLPTYPWQRQYWPDIEPDSRRHAAADPTQGFYRVDMWPEIPRSLOKS	2436
Db	1297	VFGSIGGSYAAGNAFLDALDLRRTOGLAALSANGLWAEAGMGSOAQORREHEASGIWA	QY	36	-----35	
QY	24	-----23	Db	2437	BEASRGSMVLADKGGVGEAVAAALSTRGLPCVVLHAPETSATAELVTEAAGGRSDQV	2496
Db	1357	MPTSRALAAMEWLLGTRATORVVIOMDAHAGAAPRDASGRFMDRLVTATKEASSAVP	QY	36	-----39	
QY	24	-----23	Db	2497	VLYLWGLDVAVVGAEASIDEIGDATRRATAPVLGLARFLSTVSCSPRLVVVTRGACIVGDE	2556
Db	1417	AVERNNASVETRSALYELVRGVVAGVMGFTDQGLDVRRGFABEQGLDSLMAVEIRKEL	QY	40	-----39	
QY	24	-----23	Db	2557	PAIAPCOALWGMGRVAALHHPGAWGLVDLDPASPPQASPIDGEMLTLLSQTEDQ	2616
Db	1477	QGLGMLPSATLAFDHTPTVERLVEYLLSOALELQDRDTRSVRLPATEDPIAVGAACRF	QY	40	-----39	
QY	24	-----EDTERV-----29	Db	2617	LAFRHRRHAARLVAAPPOQAAPVLSAEASYLVTTGGLGLIIVAQWLVELGARHLVL	2676
Db	1537	PGVEDESWQLLTGCVVSTEVPADRWNGADGRVPGSGEAOQRTYVPRGGFLREVETF	QY	40	-----39	
QY	30	-----29	Db	2677	TSRGLPDRQAWCEQOPEIRARIATAVEALEARGARTVAAVDVADVEPNTALVSSVEPP	2736
Db	1597	DAAPPHISPREAMSLDPOORLLLEVSWEAIERAGODPSALRESPTCVFVGAGPNEYAERV	QY	40	-----39	
QY	30	-----29	Db	2737	LRGVVHAAGVSVMRPLAETDETLLSVLRPKVAGSMLLHRLHGRPLDLFLVLFSSGAAYW	2796
Db	1657	QELADEAAGLYSGTGNMLSYAAGRLSFFLGLHPTLAVDTACSSSLVALHLCQSLRGE	QY	40	-----39	
QY	30	-----29	Db	2797	GSHSQGAYAAANAFDGLAHLRRSQSLPALSVANGLWAEGBMADAENHARLSDIGVLPM	2856
Db	1717	CDQALYGVNMLLSPKTFALLSRMHALSPPGRCCKTFSADADGYARAEBCAVVVLKRLSDA	QY	40	-----39	
QY	30	-----29	Db	2857	TSAALSALQRLVETGAARTVTRMDWARPAPVYTARGRRNLLSALVAGRDIAPSPPPAA	2916
Db	1777	QRDRDPILAVIRGTAINHDGPSSGLTVPSGPAQEALLRQALAHAGVVPADVDFVECHGTG	QY	40	-----39	
QY	30	-----29	Db	2917	TRNWRGLSVAEARVALHEIIVHGAVARVLGFLDPSALDPGMGFNEQGLDSLUMAVEIRNLLQ	2976
Db	1837	TALGDPFIEVRAUSDVYGOAPADRLPLILGAANKANLGHMEPAAGLAGLLKAVIALGQEQIP	QY	40	-----39	
QY	30	-----29	Db	2977	AELDVRLLSTTLAFDHPVTQORLVEHLLVDVLKLEDRSDTOHVRSLASDEPIAIVGAACRFP	3036
Db	1897	AQPELGELNPLLPWEALPVAVARAAPVWPRTDRPRFAGVSSFGMSGTNAHVVLBEAPAVE	QY	40	-----39	
QY	30	-----29				

```
Db 3037 GGVEDLESYWQLLAEGVYVAEVPADRWDAADWDYDPDEIPGRTYVTKGAFLRDLQRLDA 3096
Qy 40 ----- 39
Db 3097 TFFRISPREAMSLDPQORLLLEVSWEALESAGIAPDTLRSPTGTVFGAGNEYYYQRLR 3156
Qy 40 ----- 39
Db 3157 GFTDGAAGLYGGTGMLSVTAGRLSFFLGLHGPTLAMDTACSSSLVALHLACOSLRIGEC 3216
Qy 40 ----- 39
Db 3217 DOALVGGVNVLLAPETFVLLSRMRALSPDGRCKTFSADADGYARGEGCAVVVLKRLRDAQ 3276
Qy 40 ----- 42
Db 3277 RAGDSIILALIRGSANVHDGSSGLTVPGPAQQLLRQALSQAGVSPVDVFVECHGTGT 3336
Qy 43 --Dp----- 44
Db 3337 ALGDPIEQALSEVYGPGRSGDRPLVLGAAKANVAHLEAASGLASLLKAVLALRHEQIPA 3396
Qy 45 ----- 44
Db 3397 QPELGELNPHLPWNTLPVAVPRKAVPNWRCARPRRAGVSAGFSLGNTNVHVLEEAPVEP 3456
Qy 45 ----- 44
Db 3457 APAAPARPVELVLSAKSAAALDAAAARLSAHLAHPESLGDVAFSLATTRSPMEHRLA 3516
Qy 45 ----- 44
Db 3517 IATTSREALGALDAAAQOKTPOGAVRGKAVSSRGKLAFLFTGOGAQMPGMGRGLYETWP 3576
Qy 45 ----- 44
Db 3577 AFREAFDRCVALFDREIDQPLREVWMAAPGLAQAAARLDQTAQAPALFALEYALAAWLRS 3636
Qy 45 ----- 44
Db 3637 WGVPEPHVLGHSIGELVAACVAGVFSLEDVAVRLVAARGRLMQALPAGGAMVAIAASEAEV 3696
Qy 45 ----- 50
Db 3697 AASVAPHAATVSIAAVNGPDVAVIAGAEVQVLALGATTAARGIRTKRLAVSHAFHSP LMD 3756
Qy 51 ----- 50
Db 3757 PMLEDFORVAATIAYRAPDRPVVSNVTGHVAGPEIATPEYVWVRHRSVAVRFGDGAKALHA 3816
Qy 51 ----- 50
Db 3817 AGAATFVEVGPKPVLGLLPACLGEADAVLVPSLRADRSECEVVLAAALGAWYAWGGALDW 3876
Qy 51 ----- 52
Db 3877 KGVFPDGAARRVALPMYPWQRRHWNLDLTPRSAAAPAGIAGRPLAGVGLCMPGAVLHHVLS 3936
Qy 53 ----- 55
Db 3937 IGPRHQPLGDHLVFGKVVVPGCAPHVAVILSIAAERWPERAIETGTVEFLKAIAMEPDQE 3996
Qy 56 ----- 55
Db 3997 VELHAVLTPEAAGDGLFELATLAA PETERRWTHARGRVQPTDGAPGALPRLEVLEDRA 4056
Qy 56 ----- 55
Db 4057 IQPLDPAGFLDRLSAVRIGMWPLWRWLQDGRVGDSEASLATLVPTYNAHDVAPLHPILLD 4116
Qy 56 ----- 55
Db 4117 NGFAVLSLSTRSEDDGTPLPFAVERVWRWRAPVGRVRCGVPGRSQAGVSSFFLVDE 4176

Qy 56 ----- 55
Db 4177 TGEVVAEVEGFCRRAPREVFLRQESGASTAALYRLDWPEAPLPDAPAEIRTEESWVVVNA 4236
Qy 56 ----- 55
Db 4237 PGSEMAAALATRLNRCVLAEPKLEAALAGVSPAGVICLWEAGAHEEAPAAQORVATEGL 4296
Qy 56 ----- 55
Db 4297 SVVQALRDRAVRLWVTHMGAVABAGERVQVATAPVWGLGRTVMQERPELSCTLVLDLEPE 4356
Qy 56 ----- 65
Db 4357 ADAARSADVLLRELGRADDETQVAFRSGKRVARLVKATTPGGLLVDAESYRLAQOKG 4416
Qy 66 ----- 65
Db 4417 TLDQLRLAPQORRAPGPGCEVEIKVTASGLNFRTVLAVLGMYPGDAGPMGDCAGVATAVG 4476
Qy 66 ----- 65
Db 4477 QGVRHVAVGDVMTLGTLRHFVTVDARLVVRQPAGLTPAQAAATVPVAFLTAMLALHDLGN 4536
Qy 66 ----- 65
Db 4537 LRRGERVLIHAAAGGVMAAJOIARWIGAEVFATASPKWAAVQAMGVPRTHIASRTILE 4596
Qy 66 ----- 65
Db 4597 FAETFRQVTGCRGVVNLNLAGFVDASLSLLSTGGRFLEMKGKTDIRDRAAAVAAAHPGV 4656
Qy 66 ----- 65
Db 4657 RYRVFOILELAPDRTRTEILERVVEGFAAGHLRALPVHAFITKAEAAFRFMAQARHOGKV 4716
Qy 66 ----- 65
Db 4717 VLLPAPSAAPLAPTGTVLLTGGLGALGHVARWLAQOQGVPHVLTGRRGLDTPGAAKAVA 4776
Qy 66 ----- 65
Db 4777 ETEALGARVTIAASDVADRNALEAVLOAI PAEMPLQGVIIHAAGALDDGVLDDEOTTDRFSR 4836
Qy 66 --APNIT----- 70
Db 4837 VLPKVTGAWNLHELTAGNDLAFVLFSSMSGLLGSAGQSNYAAANTFLDALAAHRAEG 4896
Qy 71 -----ANLTSSL----- 77
Db 4897 LAAQSLANGPWSGDMGMAAGLSAALQARLARHGMGALS PAOCTALLGOALARPETOLGAMS 4956
Qy 78 ----- 77
Db 4957 LDVRAASQASGAAPPVVRALVRAEARHTAAGAQAALALGALPEARRADEVRKVQAE 5016
Qy 78 ----- 77
Db 5017 IARVLSWSAASAVPVDRLPSDLGLDSUTAVELRNVLGQRVGATLPATLAFDHTVDALTR 5076
Qy 78 ----- 77
Db 5077 WLDDKVLAVAEPSVSSAKSPQVALDEPIAIIIGICFRPGVADPSFWRLLESGSDAVV 5136
Qy 78 ----- 77
Db 5137 EVPHERWDIOAFYDDPDPDVRGKMTTRFGFLSDIDRDPAPFFGISPREATTMDPQORLLL 5196
Qy 78 ----- 77
Db 5197 ETSWEAFERAGILPERLMGSDTGVFVGLFYQBYAALAGGIEAPDGYLGTGTASVAGRI 5256
```

QY 78 ----- 77
Db 5257 SYVLGKPSLTVDTCSSSLVAVHLACQALRRGECVALGAGVALMLTPATFVFSRLR 5316
QY 78 ----- 77
Db 5317 GLAPDRCKSFAADGVGSEGCAMLLKPLRDAQRDGDPI LAVIRGTAVNQDGRSNGL 5376
QY 78 ----- 77
Db 5377 TAPNGSQOEVRRALEQAGLAPADVSYVECHGTGTLGDPTEVOALGAVLAQGRPSDRP 5436
QY 78 ----- 77
Db 5437 LVIGSVKSNIGHTQAAAGVAGV KVALALERGLIPRSLHFDAPNPHIPWSELAVQVAAPK 5496
QY 78 ----- 77
Db 5497 VENTRNGVPRRAGVSSFGVSGTNAHVLEAPAAAFAPAAARSAEFLVLSAKSAALDAQ 5556
QY 78 ----- 77
Db 5557 AARLSAHVVAHPGLGLDPLAFSLATRSPMTYRLAVAATSREALSAAALDTAAQOAPPA 5616
QY 78 ----- 84
Db 5617 ARGHASTGAPKVPVFPQGSQWLGMOKLLSEBPFRDALSACDRAIOAEAGWSLLAE 5676
QY 85 ----- 84
Db 5677 LAADETTSQLRIDVWPALFAIEVALSALWRSWGPEDVAVGHSGEVAHAHVAGALS 5736
QY 85 ----- 84
Db 5737 EDAVAICRRSILLRRISGQEMAVVELSLAEAEALLGYEDRLSVAVSNSRSTVLAGE 5796
QY 85 ----- 84
Db 5797 PAALAEVLAILAAKGVFCRRVKVDVAGHSPQIDPLRDELLAALGELEPRQATVSMRSTVT 5856
QY 85 ----- 84
Db 5857 STIMAGPELVASYWADNVQVRFPAEAVQSLMEDGHGLFVEMSPHPILTTSVEEIRATK 5916
QY 85 ----- 84
Db 5917 REGVAVGLRRGODERLSMLLEALCALWVHQAVGWERLFSAGCAGLRVPLTPYMQRER 5976
QY 85 ----- 90
Db 5977 YWVDAPTGAAGSRFAHAGSHPLLGEMQTLSTQSTRVWETTLDLKRLPWLGDHRVQGA 6036
QY 91 ----- 94
Db 6037 VVFPGAYLEMALSSGAELGDPLOVSDVVLAEALAFADDTPAAVQVQWATEERPCRLQF 6096
QY 95 ----- 97
Db 6097 HVASRVPGHGGAFRSHARGVLRQIERAEVPAFLDLAALRLARLQASAPAAATYAAAEWG 6156
QY 98 ----- 97
Db 6157 LEYGAFOGLVELWRGEALGRVRLPEAGSPAACRLHPALLDACFHVSSAFADRGEAT 6216
QY 98 ----- 97
Db 6217 PWVPVEIGSLRFPQSGELWCHARSVSHGKPTDPRSTDFWVVDSTCAIVAEISGLVAQ 6276
QY 98 ----- 97
Db 6277 RLAGGVRRREDDWFEPAPWEPTAVPGSEVMAGRMLLIGSGGLGAALHSALTBAHSV 6336
QY 98 ----- AD 99

Db 6337 HATGRGTSAGLQALLTASFDGOAPTSVVHLGSLDERGVLDADAPFDADALEESLVRGCD 6396
QY 100 LVLVY-----ITR----- 106
Db 6397 SVLWTVQAVAGAFRDPRLMLVTRGAQAIGAGDVSAQAPLLGLGRVIALEHAELRCAR 6456
QY 107 ----- 106
Db 6457 IDLDPARRGEVDELLEALLADDAEEVAFRGERRVARLVRRLPETDCREKIEPAERGP 6516
QY 107 FOLEL----- 111
Db 6517 FRLEIDGSGVLDDLVLRATERPPGPGVEIATVEAAGLNFLDVVRAMGIYPGPGDGPVAL 6576
QY 112 ----- 111
Db 6577 GAECGRIVAMGEGVESURIQGVVAVAPFSFGTHVTIDARMLAPRAALTAQAALPV 6636
QY 112 ----- 111
Db 6637 AFMTAWYGLVHLGRLRAGERVLIHSATGGTGLAAVOIARHLGAEIFATAGTPEKRAWLRE 6696
QY 112 -----PDGN----- 115
Db 6697 OGIAHVMDSRSLDFAEQVLAATKGEVDVVLNSLSGAAIDASLSLTPDGRFFIELGKTDI 6756
QY 116 ----- 115
Db 6757 YADRSGLAHPKRSLSYSAVDLAGLAVRRPERRVAALLAEVVDLLARGALQPLPVEIPPLS 6816
QY 116 ----- 115
Db 6817 RAADAFKMAQAQHLGKLVLALEDPPDVRIRVPGESGVAIRADGAYLVLTGGLGGLSVAG 6876
QY 116 -----ROV----- 118
Db 6877 WLABOGAGHLVLVGRSAGVSAEQOTAVAALAEHAGARTVARADVADRAQMERILREVTA 6936
QY 119 -----RGVTQLGG----- 126
Db 6937 GMPLRGVVAAGLDDGLLMOQTPARFRAVMAKPVKRGALHLHALTREAPLSFFVLVYASGA 6996
QY 127 -----ACSPWMSCL----- 135
Db 6997 GLLGSPGQGYAAANTFLDALAHRRAOGLPALSIDWGLFADVGLAAGQONRGARLVTRG 7056
QY 136 -----ITEDTG-----FD----- 143
Db 7057 TRSLTTPDEGLWALERLLDGDRTQAGVMPDVRQWVEFYPAASRRSLRMTARRVASGR 7116
QY 144 ----- 143
Db 7117 LAGDRDLERLATAEAGARAGMLQEVVRAQVSOVLRLSEGKLDVDAPLTSLGMDSLMGL 7176
QY 144 -----LGVTI 148
Db 7177 LNRNIEAVLGITW 7189

RESULT 9

US-09-568-480-5
; Sequence 5, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-5

Query Match 27.0%; Score 209; DB 4; Length 7257;
Best Local Similarity 1.5%; Pred. No. 3.1;
Matches 93; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;
QY 1 AAGG----- 4
|||
Db 877 AAGGSWPGVFTAGRRVPLTPYPMORQRYWIEAPAEGLGATAADALAQWFYRVDPWPM 936
QY 5 -----ILH----- 7
|||
Db 937 PRSSVDSRRARSGWLVLADRGVGEAAAAALSSQSCCAVLHAPAEASAVAEQVTOALG 996
QY 8 ----- 7
|||
Db 997 GRNDWQGVLYLWGLDAWEAGASAEVAKVTHLAAAPVLALIQALGTGPRSLWVTRG 1056
QY 8 ----- 7
|||
Db 1057 ACTVGEPDAAPCQAAALWGMGRVAALHHPGSGWGLVDLPRESPTVEALVAELLSDAE 1116
QY 8 ----- 7
|||
Db 1117 DQAFRQGRRAARLVAAPPEGNAAPVLSAEGSVLTGGLGALLVARWLVERGAGHL 1176
QY 8 -----LE-----LLVAVG 15
|||
Db 1177 VLISRHLDPREWRDQPEVRARIAIEALEAQAGRTVAADVADAEAGMAALLAAVE 1236
QY 16 P-----DVFAQHQ----- 23
|||
Db 1237 PPLRGVHAAGLLDDGLLAHQDAGRLARVLRPKVEGANVHLTLTREQPLDLFVLFSASG 1296
QY 24 ----- 23
|||
Db 1297 VFGSIGQSYAAGNAFLDALADLRRTQGLAALSIAWGLWAEAGMGSAQRREHESGIWA 1356
QY 24 ----- 23
|||
Db 1357 MPTSRALAAEWLLGTRATORVVIQMDWAHAGAAPDASRGRFWDRLVTATKEASSAVP 1416
QY 24 ----- 23
|||
Db 1417 AVERWNASVETRYSALVELYRGVVGVGMFTDQGLDVRRAEQGLSLMAVEIRKRL 1476
QY 24 ----- 23
|||
Db 1477 QGELGMLPLATLAFDHTPVERLYLLSQALEQDRTDVRSLPATEDPIAIVGAACRF 1536
QY 24 -----EDTERY----- 29
|||
Db 1537 PGGVEDLESYWQLLTGEGVWVSTEVPADRWNGADGRVPGSGEAOQTYYPRGGFLREVTF 1596
QY 30 ----- 29
|||
Db 1597 DAAPFHSPREMSLDPOQRLLLEVSWEATERAGODPSALRESPTGVFGAGFNEYAERV 1656
QY 30 ----- 29
|||
Db 1657 QELADEAAGLYSGTGNMLSVAAGRLSFFLGLHPTLAVDTACSSSLVALHGCQSLRGE 1716
QY 30 ----- 29
|||

Db 1717 CDQALVGVGNMLLSPKTFALLSRMHALSPGCRCKTFPSADADGYARAGCAVVVLKRLSDA 1776
QY 30 ----- 29
|||
Db 1777 QRDRDPILAVIRGTAINHDGPSSGLTVPSPGPAOBALLRQALAHAGVVPADVDFVECHGTG 1836
QY 30 ----- 29
|||
Db 1837 TALGDPIEVRALSDVYGOARPADRPRLILGAAKANLGHMEPAAGLAGLLKAVLALGOEIP 1896
QY 30 ----- 29
|||
Db 1897 AOPELGELNPLLPWEALPVAVARAAVWPRTDRPRFAGVSSFGMSGTNAHVLEEAPAVE 1956
QY 30 ----- 29
|||
Db 1957 LMPAAPERSAELLVLSCKSEGLADAQAARLUREHLDHPELGLGDVAFSLATTRSAMNHL 2016
QY 30 ----- 29
|||
Db 2017 AVAVTSREGLLAALSAVAQOTPPGAARCIASSSRCKLAFLTGQCAQTTPGMRGLCAAW 2076
QY 30 ----- 29
|||
Db 2077 PAFREAPDRCVALLFDRELDRLPREVMABPGSAESILLDDTQFTOPALFTVEYALTALWR 2136
QY 30 ----- 29
|||
Db 2137 SWGVEPELVAGHSAGELVAACVAGVPSLEDGVLVAARGRLMOGLSAGGMVSLGAPEAE 2196
QY 30 ----- 29
|||
Db 2197 VAAVAHAPHAASVIAAVNGPEQVVIAGVQAVQAIAGFAARGARTKRLHVSHASHSPLM 2256
QY 30 -----VLTNLN----- 35
|||
Db 2257 EPMLEFGRVAASVTRYRPSVLSVNSLKGKVADELSPGYVVRHVREAVRPAFCVKALH 2316
QY 36 ----- 35
|||
Db 2317 EAGAGTFVEVGPKPTLLGLLPACLPEABPTLLASURAGREAAAGVLEALGRMAAGSVS 2376
QY 36 ----- 35
|||
Db 2377 WPGVFTAGRRVPLPTYWQQRWYWDIEPDSRRHAAADPTQGWFYRVDMPEIPRSLQKS 2436
QY 36 ----- 35
|||
Db 2437 BEASRGMLVLADKGVGEAAALSTRGLPCVVLHAPAEATSATAELVTEAGGRSDWQV 2496
QY 36 -----TGAE----- 39
|||
Db 2497 VLYLWGLDADVVGAEASIDEIGDATRRATAPVLGLARELSTVSCSPRLWVVRGACIVGDE 2556
QY 40 ----- 39
|||
Db 2557 PAIAPCOAALWGMGRVAALHHPGANGGLVDLDPASPQASPIDGEMLVTELLSQETEDQ 2616
QY 40 ----- 39
|||
Db 2617 LAFRRHRHAARLVAAPPOQAAAPVLSABASVYLTGGLGGLLIVAQMLVELGARHLVL 2676
QY 40 ----- 39
|||
Db 2677 TSRRGLPDRQACEQPPEIRARIAAVEALEARGARTVAADVADVEPMALVSSVEPP 2736
QY 40 ----- 39
|||
Db 2737 LRGVHAAGVSVMRPLAETDETLLSVLRPKVAGSMLLHRLHGRPLDLFVLFSSGAANW 2796
QY 40 ----- 39
|||

```
Db 2797 GSHSOGAYAAANAFDGLAHLRRSQSLPALSVAWGLWAEGMADAHAHRLSDIGVLPMS 2856
QY 40 -----HLV----- 53
Db 2857 TSAALSALORLVETGAQRTVTRMDWARFAPVYTARGRRNLLSALVAGRDIIAPSPAAA 2916
QY 40 ----- 39
Db 2917 TRNWRGLSVAEARVALHEIVHGAVARVLGFLDPSALDPGMGFNEQGLDSLMAVEIRNLQ 2976
QY 40 ----- 39
Db 2977 AELDVLSTTLAFDHTVQRLVEHLLVDVLKLEDRSDTQHVRSLASDEPTAIVGAACRFP 3036
QY 40 ----- 39
Db 3037 GGVEDLESYWOLLAEGVVVAEVPADRMDAADWYDPDPEIPGRYTVTKGAFRLDQLRLDA 3096
QY 40 ----- 39
Db 3097 TFRISPREAMSLDPQORLLLEVSWEALESAGIAPOTLRDSPGCVFVGAGPNYYTORLR 3156
QY 40 ----- 39
Db 3157 GFTDGAAGLYGGTNMLSVTAGRLSFFLGLHPTLAMDTACSSSLVALHLACQSLRGEC 3216
QY 40 ----- 39
Db 3217 DQALVGGVNVLLAPETFVLLSRMRALSPDGRCKTFSDADGYARGCCAVVVLKRLRDAQ 3276
QY 40 -----LLR----- 42
Db 3277 RAGDSLILALIRGSVNHGDPSSGLTVPNGPAQALLRQALSQAGVSPVDVFVECHGTGT 3336
QY 43 ---DP----- 44
Db 3337 ALGDPIEVOALSEVYGRSGDRPLVLGAAKANVAHLEAASGLASLAKVALLRHQIPA 3396
QY 45 ----- 44
Db 3397 QPELGELNPHLPWNTLPVAVPRKAVPWGRGARPRRAGVSAFGLSGTNVHVLEEAPEVEP 3456
QY 45 ----- 44
Db 3457 APAAPARVELVLSAKSAAALDAAAARLSAHLSAHPELSLGDAVFSLATTRSPMEHRLA 3516
QY 45 ----- 44
Db 3517 IATTSREALRGALDAAAQOKTPOGAVRGKAVSRGKLAFLEFTQGGAQMPGMRGLYETWP 3576
QY 45 ----- 44
Db 3577 AFREAFDRCVALFDREIDQPLREVWMAAPGLAQOARLDOTAYAQPALFALEYALAAALWRS 3636
QY 45 ----- 44
Db 3637 WGVEPHVLLGHSIGELVAACVAGVFSLEDAVRLVAARGRLMOALPAGGAMVAIAASEAEV 3696
QY 45 -----SLGQF----- 50
Db 3697 AASVAPHAATVSTAAVNGPDVAVVIAGAEVOVLALGATFAARGIRTKRLAVSHAFHSPLMD 3756
QY 51 ----- 50
Db 3757 PMLEDFORVAATIAYRAPDRPVVSNVTGHVAGPEIATPEYVVRHVRSVAVRFGDGAKALHA 3816
QY 51 ----- 50
Db 3817 AGAATEVEYGPKEVLLGLLPACLIGEADAVLPSLRADRSECEVLLAALGAWYAWGALDW 3876
QY 51 -----RV----- 52
Db 3877 KGVFPDGAARRVALPMYPWQRRHRHMDLTPRSAAPAGIAGRWPLAGVGLCMCPGAVLHHVLS 3936
QY 53 -----HLV----- 55
Db 3937 IGRPHQFPGDLHLVFGKVVVPGAFHVAVILSIAAERWPERAIETLGVFELKAIAMEPDOE 3996
QY 56 ----- 55
Db 3997 VELHAVLTPEAAGDGYLPELATLAAPETERRTWTHARGRVQPTDGPAGALPRLEVLEORA 4056
QY 56 ----- 55
Db 4057 IQPLDFAGFLDRLSAVRIGWGPLMRWLQDGRVGDDEASLATLVPTYPNADHAPLHPILLD 4116
QY 56 ----- 55
Db 4117 NGFAVSLLSSTRSEPEDDGTPLPFAVERVRWRAPVRGVRGCVPRSOAFGVSSFVLVDE 4176
QY 56 ----- 55
Db 4177 TGEVVAEVEGFVCRRAPREVFLROESGASTAALYRLDWPPEAPLPDAPAERIEESWVVVAA 4236
QY 56 ----- 55
Db 4237 PGSEMAAALATRLNRCVLAEPKGLEAALAGVSPAGVICLWEAGAHEEAPAAAQRVATEGL 4296
QY 56 ----- 55
Db 4297 SVVQALDRADVRLMWVTMGAVAEGERVQVATAPVMGLGRTVMQERPELSCTLVDLEPE 4356
QY 56 -----KMWILTPEP----- 65
Db 4357 ADAARSADVLLRELGRADDETQVAFRSQKRRVARLVKATTPGEGLLVPPDAESYRLEAGOKG 4416
QY 66 ----- 65
Db 4417 TLDQLRLAPORRAPGGEVEIKVTASGLNPTVLAVLGMYPGDAGPMGDCAGVATAVG 4476
QY 66 ----- 65
Db 4477 QGVRHVAVGDVMTLGLTHRFVTVDARLVVRQAGLTPAQOATVPVAPLTAWLALHDLGN 4536
QY 66 ----- 65
Db 4537 LRRGERVLIHAAAGGVGMAAVQIARWIGAEVFATASPSKWAAVQAMGVPRTHIASRTLE 4596
QY 66 ----- 65
Db 4597 PAETFRQVTGGRVDVVLNALAGEFVDASLSLLSTGGFLEMGKTDIRDRAAVAAAHPGV 4656
QY 66 ----- 65
Db 4657 RYRVFDILELAPDRTREILERVVEGFAAGHLRALPVHAFITKAEAAFRFMAQARHOKV 4716
QY 66 ----- 65
Db 4717 VLLPAPSAAPLAPGTGTLLTGGLGALGLHVARMLAQQGVPHMVLTRRRGLDTPGAAKAVA 4776
QY 66 ----- 65
Db 4777 EIEALGARVTITAAASDVADRNALEAVLOAIIPAEMWLOGVIHAAGALDDGVLDDEOTTDRFSR 4836
QY 66 ---APNIT----- 70
Db 4837 VLAPKVTGANNLHELHETAGNDLAFVFLPSSMSGLLSAGOSNYAAANTFLDLAAHRAEG 4896
QY 71 -----ANLTSSL----- 77
Db 4897 LAAQSLAWGPNWSDGMAAGLSAALQARLARHGMGALSQAQGTALLGQALARPETQLGAMS 4956
QY 78 ----- 77
Db 4957 LDVRAASQASGAAVPPVWRALVRAEARHTAAGAOGAALARLGAALPEARRADEVRKVVQAE 5016
```

```
QY 78 ----- 77
Db 5017 IARVLSAASAVPVDRPLSDGLDLSLTAVELRNVLGQVATLPAFLDHPDPTDALTR 5076
QY 78 ----- 77
Db 5077 WLLDKVLAVAPSVSSAKSSPQVALDEPIAIIIGICRPPGGVADPESFWRLLEBGSDAVV 5136
QY 78 ----- 77
Db 5137 EPPERWDIDAFYDDPDVRCMTTRFCGFLSDIDRDPAPFGISPREATTMPQORLLL 5196
QY 78 ----- 77
Db 5197 ETSWEAFERAGILPERLMGSDTVFVGLFYOEYAAAGGIEAFDGYLGTGTASVAGRI 5256
QY 78 ----- 77
Db 5257 SYVILGKPSLTVDTACSSSLVAVHLACQALRRGECSVALAGGVALLMTPATVEFSRLR 5316
QY 78 ----- 77
Db 5317 GLAPDGRCKSFAAADGVHSEGCAMLLKPLRDAQRDGDPILAVIRGTAVNQDGRSNGL 5376
QY 78 ----- 77
Db 5377 TAPNGSSQEVIRRALEQAGLADVSVECHGTGTTGLDPIEQALGAVLAQGRPSDRP 5436
QY 78 ----- 77
Db 5437 LVIGSVKSNIGHTQAAAGVAGVIKVALALERGULIPRSLHFDAPNHPIMSELAVQVAAKP 5496
QY 78 ----- 77
Db 5497 VENTRNGVPRRAGVSFGVSGTNAHVHLEAPAAAPAAARSaelFVLSAKSAAALDAQ 5556
QY 78 ----- 77
Db 5557 AARLSAHVAHPGLGDLAFSLATTRSPMTYRLAVAATSREALSAALDTAAQGPAA 5616
QY 78 ----- 84
Db 5617 ARCHASTGNAPKVVVFPPGGGQWLGOKLLSEEPVFRDALUSACDRAIOAEAGWSLLAE 5676
QY 85 ----- 84
Db 5677 LAADETTSQGRIDVVQPALFAIEVALSALWRSWGVEPDVVGHSWGEVAHAHVAGALS 5736
QY 85 ----- 84
Db 5737 EDVAVALICRRSILLRRISQGGEMAVVELSLAEAEALLGYEDRLSVAVSNSPRSTVLAGE 5796
QY 85 ----- 84
Db 5797 PAALAEVLAILAAGVFCRRVKVDVASHSPQIDPLRDELLAALGELEPRQATVSMRSTVT 5856
QY 85 ----- 84
Db 5857 STIMAGPELVASYADNVRQPVFAEAVQSLMEDGHGLFVEMSPHILTTTSVEIRRAK 5916
QY 85 ----- 84
Db 5917 REGVAVGSLRRCODERLSKLEALGALVHGQAVGWERLFSAGGAGLRRVPLTPYPMQORER 5976
QY 85 ----- 90
Db 5977 YWVDAPTGAAGGRFAHAGSHPLLGEMQTLSTQSTRVWETTLDLKELPWLGDRHVOGA 6036
QY 91 ----- 94
Db 6037 VVFGCAVLEMASSGAELGDPLOVSDVVLAELAFADDTPAAVQVMATEERPGRLOF 6096
QY 95 ----- PGH ----- 97

|||
Db 6097 HVASRVPGGGAARSHARGVLRQIERAEVPARLDLAALARLOASAPAAATYAALAEWG 6156
QY 98 ----- 97
Db 6157 LEYGPAFOGLVELWRGEGBALGRVRLPEAAGSPAACRLHPALLDACFHVSSAFADRGBAT 6216
QY 98 ----- 97
Db 6217 PMWVPEIGSLRWFQPSGELWCHARSVSHGKPTPDRRSTDFMVVDSTCAIVAEISGLVAQ 6276
QY 98 ----- 97
Db 6277 RLAGVRRREDDWFMPEWPEPTAVPGSEVMAGRMLLIGSGGLGALHLSALTAEAGHSV 6336
QY 98 ----- AD 99
Db 6337 HATGRGTSAGLQALLTASFDGOAPTSMVHLGSLDERGVLDADAPFDADALEESLVRCGD 6396
QY 100 LVLY ----- ITR ----- 106
|||
Db 6397 SVLWTVQAVAGAGFRDPPRLWLVRGAQAIGAGDVSVAQAPLLGLGRVIALEHAELRCAR 6456
QY 107 ----- 106
Db 6457 IDLDPARRDGEVDELLAEALLADDAEEVAFRGGERRVARLVRRLPETDCREKIEPAGRP 6516
QY 107 FDLEL ----- 111
|||
Db 6517 FRLEIDGSGVLDLVLRTERRPPGPCEVEIAVEAAGLNFLDMVRANGIYPPGCGDPVAL 6576
QY 112 ----- 111
Db 6577 GABCSGRIVAMGEVSESLRIGODVAVAPFSFGTHVTIDARMLAPRPAALTAQAAALPV 6636
QY 112 ----- 111
Db 6637 AFMTAWYGLVHLGRLAGERVLIHSATGGTGLAAVOIARHLGAEIFATAGTPEKRAWLRE 6696
QY 112 ----- PGN ----- 115
|||
Db 6697 QGIAHVMSRSLDFAEQVLAATKGEQDVVNLNSLGAADSLTSLVDPGRFIELKCTDI 6756
QY 116 ----- 115
Db 6757 YADRSGLAHFRKSLSYSADVLAGLAVRRPERVAALLAEWOLLARGALQPLPFIPLS 6816
QY 116 ----- 115
Db 6817 RAADAFKVAQAQHLGKLVLALEDPDVIRIVPGESGVAIRADGAYLVTTGGGLGLSVAG 6876
QY 116 ----- ROV ----- 118
|||
Db 6877 WLAEOGAGHLVLVGRSGAVSABEQQTAVAALAEHARGVTVARADVADRAQMERILREVTA 6936
QY 119 ----- RGVTOGG ----- 126
|||
Db 6937 GMPLRGVVHAAGILDDGLLMQOTPARFRAVMAKPVKRGALHHLHALTREATPLSFVLYASGA 6996
QY 127 ----- ACSPTWSCL ----- 135
|||
Db 6997 GLLGSPGQGNAAAANTFLDALAHHRRAOGLPALSIDMGLPADVGLAAGQONRGARLVTRG 7056
QY 136 ----- ITEDTG ----- FD ----- 143
|||
Db 7057 TRSLTPDEGLWALERLLDGRDTOAGWMPFDVRQWVEFYPAASASSRRLSRLMTARRVASGR 7116
QY 144 ----- 143
Db 7117 LAGDRDLERLATAEAGARAGMLQEVVRAQVSQVLRLSEGLDLDVDAPLTSLGMSLMLGLE 7176
QY 144 ----- LGVTI 148
|||
```

Db 7177 LRNRIEAVLGITM 7189

RESULT 10

US-09-568-486-5

; Sequence 5, Application US/09568486

; Patent No. 6355459

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/568,486

; CURRENT FILING DATE: 2000-05-10

; PRIOR APPLICATION NUMBER: 09/335,409

; PRIOR FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 7257

; TYPE: PRT

; ORGANISM: Sorangium cellulosum

US-09-568-486-5

Query Match 27.0%; Score 209; DB 4; Length 7257;

Best Local Similarity 1.5%; Pred. No. 3.1;

Matches 93; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;

Qy	1	AAGG-----	4
Db	877	AAGGSWSGVFTAGRRVPLTPYQRYWIEAPAEGLGATAADALAQWFYRVWDNDEM	936
Qy	5	-----ILH-----	7
Db	937	PRSSVDSRRASGWLVLADRGVGEAAALSSQGCSCAVLHAPAEASAVAEQVTOALG	996
Qy	8	-----	7
Db	997	GRNDWQGVLYLWGLDAVVEAGASAEVAKVTHLAAAPVLALIOALGTGPRSPRLMTVTRG	1056
Qy	8	-----	7
Db	1057	ACTVGGEPPDAAPCOALMGWRVALEHPGSGWGLVDLPDESPTEVEALVAELLSPDAE	1116
Qy	8	-----	7
Db	1117	DQLAFQRRRAARLVAAPEEGNAAPVLSAEGSYLVTTGGLGALGLVARMLVERGAGHL	1176
Qy	8	-----LE-----	15
Db	1177	VLISRHGLPDREWGRRDQPEVVRARIAATEALEAQGVTVAAVDVADAEAGMAALLAAVE	1236
Qy	16	P-----DVFOAHQ-----	23
Db	1237	PPLRGVVAAGLDDGLAHQDAGRLARVLRPKVEGAWVLHTLTREQPLDLFLVLFSSAG	1296
Qy	24	-----	23
Db	1297	VFGSIGQSYAAGNAFLDALADLRRTOGLAALSIAWLWAEAGMGMSQAORREHEASGIWA	1356
Qy	24	-----	23
Db	1357	MPTSRALAAWELLGTRATORVVQMDWAHAGAAPRDAASGRFRWDRLVATKEASSAVP	1416
Qy	24	-----	23
Db	1417	AVERWRNASVETRSALVELVRGVAGVNGFTDQGTLDVYRGFAEQGLDLSLMAVEIRKEL	1476
Qy	24	-----	23

Db	1477	QGEGLMPLSATLAFDHPHTVERLVEYLLSQALELQDRTDVRSLPATEDPDIATVCAACRF	1536
Qy	24	-----EDTERTY-----	29
Db	1537	PGGVEDLESYQQLLTGEGVVVSTEVPADRWNGADGRVPGSGEAOQRTYVPRGGFLREVETTF	1596
Qy	30	-----	29
Db	1597	DAAFFHISPREAMSLDPOORLLLEVSWEAIERAQDPSALRESPTGVFVGAGPNEYAERV	1656
Qy	30	-----	29
Db	1657	QELADEAAGLYSGTGNMLSVAAGRLSPFLGLHGPTLAVDTACSSSLVALHLCOSLRGE	1716
Qy	30	-----	29
Db	1717	CQALVGGVNMLLSPKTFALLSRMHALLSPGCRCKTFSADADGYARAEGCAVVLKRLSDA	1776
Qy	30	-----	29
Db	1777	QRDRDPI LAVIRGTAINHDGSPSGLTVPSGPAQBALLRQALAHAGVVPADVDFVECHGTG	1836
Qy	30	-----	29
Db	1837	TALGDPIEVRALSDVYGOARPADRPRLILGAAKANLGHMEPAAGLAGLLKAVLALGOEIP	1896
Qy	30	-----	29
Db	1897	AQPELGEINPLLPWEALPVAVARAAPWPRTRDRPRFAGSVSGMGTNAHVLEADPAVE	1956
Qy	30	-----	29
Db	1957	LWPAAPERSAELLVLSCSEGLDAQAARLREHLDMHPGLDGLDVAFLATTRSAMNHL	2016
Qy	30	-----	29
Db	2017	AVAVTSREGLLAALSAVAOQTTPGAARCIASSSRGKLAFLFTQOQAOTPCMGRGLCAAW	2076
Qy	30	-----	29
Db	2077	PAFREAFDRCVALFDRELDRLREVMMWAEFGSAESLLDQTAFTQPALFTVEYALTALWR	2136
Qy	30	-----	29
Db	2137	SWGVEPELVACHSAGELVAACVACVPSLEDGVRLVAARGRLMOGLSAGGAMVSLGAPEAE	2196
Qy	30	-----	29
Db	2197	VAAAVAPHAASVSIAAVNGPEQVVVAGVEQAVQAIAGFAARGARTKRLVSHASHSPLM	2256
Qy	30	-----VLTNLN-----	35
Db	2257	EPMLEEFGRVAASVTYRRRPSVLSVNSLGSKVWAEDELSAPGYWVRHVREARFADGVKALH	2316
Qy	36	-----	35
Db	2317	EAGAGTFVEVGPRTLLGLLPACLPEAEPTLLASLRAGREAAAGVLEALGRLWAAGGSVS	2376
Qy	36	-----	35
Db	2377	WPGVFPTAGRRVPLPTYQRYWPDIEPDSRRHAAADPTOGMFYRVDMWPEIPRSLOKS	2436
Qy	36	-----	35
Db	2437	EEASRGSLVLADKGGVGEAAVAAALSTRGLPCVVLHAPAEATSATAELVTEAAGRSNQV	2496
Qy	36	-----IGAE-----	39
Db	2497	VLYLWGLDAVGAESAIDEIGDATRRATAPVLGLARFLSTVSCSPRLVVVTRGACIVGDE	2556
Qy	40	-----	39


```
Db 2557 PAIAPCOAALWGMGRVAALHEHFGAWGLVDLDPRASPPQASPIDGEMLVTELLSQETEDQ 2616
Qy 40 -----
Db 2617 LAFRHRHRAARLVAAAPGQAAAPVLSAENSYLVTGGLGGLGLIIVAQMLVELGARHLVL 2676
Qy 40 -----
Db 2677 TSRRGLPDRQAWCEQPPEIRARIAAVEALEARGARVTVAAVDVADVEPM TALVSSVEPP 2736
Qy 40 -----
Db 2737 LRGVVHAAGVSVMRPLAETDETLLSVLRPKVAGSM LLLHLLHGRPLDLFLVFSGAAVM 2796
Qy 40 -----
Db 2797 GSHSOGAYAAANAFDLGLAHLRRSQSLPALSVAVGLWAEGMADAEAHARLSIDIGVLPMS 2856
Qy 40 -----
Db 2857 TSAALSALORLVETGAAQORTVTRMDWARFAPVYTARGRRNLLSALVAGRDIIAPSPAAA 2916
Qy 40 -----
Db 2917 TRNRGLSVAEAREVALHEIVHGAVARVLGFLDPSALDPMGFENQGLSLMAVEIRNLLQ 2976
Qy 40 -----
Db 2977 AELDVRLSTTLAFDHTVQRLVEHLLVDVLKLEDRSDTOHVRSLASDEPIALVGACRFP 3036
Qy 40 -----
Db 3037 GGVEDLESYQWLLAEGVVVSAEVPADRMDAADWDPEIPGRTYTVTKGAFLRLQLRDA 3096
Qy 40 -----
Db 3097 TFRISPREAMSLDPOORLLLEVSWEALESAGIAPDTRLDSPTGVFVAGPNEYTORLR 3156
Qy 40 -----
Db 3157 GFTDGAAGLYGTGNM LSVTAGRLSFFLGLHGPTLAMDTACSSSILVALHLACQSLRLEC 3216
Qy 40 -----
Db 3217 DOALYGGVNVLLAPETFFVLLSRMRALSPDGRCKTFSADADYARGECAVVLKRLRDAQ 3276
Qy 40 -----
Db 3277 RAGDSILIRGSVNHDPSSGLTVPNGPAQALLRQALSQAGVSPVDVDFVECHGTGT 3336
Qy 43 ---DP-
|||
Db 3337 ALGDPFIEVOALSEVYGPGRSGDRPLVLGAAKANVAHLEAASGLASLLKAVLALRHEQIPA 3396
Qy 45 -----
Db 3397 OPELGELNPLPMTNLPVAVPRKAVPNRGARPRRAGVSAGLSGNTNVHVHVEEAPEVEP 3456
Qy 45 -----
Db 3457 APAAPARVELVILSAKSAAALDAAAARLSAHL SAHP ELSLGDVAFSLATTRSPMEHRLA 3516
Qy 45 -----
Db 3517 IATTSREALCALDAAAQKQTPQAVRGKAVSSRGKLAFLFTCGGAQMPGMGRGLYETWP 3576
Qy 45 -----
Db 3577 AFREAFORCVALEFDREIDOPREVMWAA PGLAQARLDQTYAQPALPALEYALAAALWRS 3636
Qy 45 -----
Db 3637 WGVPEPHVLLGHSIGELVAACVAGVFSLEDAVRLVAARGRLMOALPAGGAWVAIAASEAEV 3696
Qy 45 -----
Qy 45 -----SLGAOF----- 50
3697 AASVAPHAATYSIAAVNGPDVAVIAGAEVQVLALGATTAARGINTKRLAVSHAFSP LMD 3756
Qy 51 -----
Qy 51 ----- 50
3757 PMLEDFORVAATIAIRAPDRPVVSNVTGHVAGPEIATPEYVVRHVRSAVRFGDGAKALHA 3816
Qy 51 -----
Qy 51 ----- 50
3817 AGAATFVEVGPKPVLLGLLPACLGADAVLVPSLRADRSCEVVLAAALGAWYAWGGALDW 3876
Qy 51 -----RV-----
|||
3877 KGVPDGGARRVALPMYPMQORERHMDLTPRSAAPAGIAGRWP LAGVGLCMPGAVLHHVLS 3936
Qy 53 -----HLV-----
|||
3937 IGPRHQPFLLGDHLVFGKVVVPGAFHVAVILSIAAERMPRAIELTGTVEFLKATAMEPDOE 3996
Qy 56 -----
3997 VELHAVITPEAAGDGYL FELATLAAPE TERTTTHARGVQPTDGPAGALPRLEVEDRA 4056
Qy 56 -----
4057 IQPLDFAGFLDLRSVRI GWPLRWLQDGRVGD EASLATLVPTYNAHDVAPLHPILLD 4116
Qy 56 -----
4117 NGFANVLLSTRSEPEDDGTPLPFAVERVRWWRAPVGRVRCGVP RSOAFVSFVLVDE 4176
Qy 56 -----
4177 TGEVVAEVEGFCRRAPREVFLRQESGASTAALYRLDWPEAPLPDAPAERIEESVWVAA 4236
Qy 56 -----
4237 PGSEMAALATRLNRCVLAEPKGLEAALAGVSPAGVICLWBAGAEHEEAPAAAQRVATEGL 4296
Qy 56 -----
4297 SVVQALDRAVRLMWVTMGAVAVEGERVQVATAPVGLGRVTVMQERPELSCTLVLEPE 4356
Qy 56 -----KXVILTEPEG-----
|||
4357 ADAARSADVLLRELGRADDETQVAFRSGKRRVARLVKATTP EGLLPDAESYRLEAQKG 4416
Qy 66 -----
4417 TLDQLRLAPAORRAPGPGOEVEIKVTASGLNFRTVLAVLGMYPGDAGPMGGDCAGVATAVG 4476
Qy 66 -----
4477 QGVRHVAVGDVMTLGT LHRFTVDARLVVRQAPAGLTPQAATVPVAF LTAWLALHDLGN 4536
Qy 66 -----
4537 LRRGERVLIHAAAGGVGNAAVQIARMI GAEFVATASPKWAAVQAMGVPRTHIASRRTLE 4596
Qy 66 -----
4597 FAETFRQVTGGRGVVDVNLNAGFEFVDASLSLLSTGGRFLEMKGKDIRDRAAVAAAHPGV 4656
Qy 66 -----
4657 RYRVFDILELAPORTREILERVVEGFAAGHLRALPVHAFATKAEAAFRMAOARHOGKV 4716
Qy 66 -----
4717 VLLPAPSAAPLAPTGTVLLTGG LUGALGLHVARWLAAOQGVPHMVL TGRRGLDTPGAARAVA 4776
```

QY	66	-----	65
Db	4777	EIEALGARVTIAASDVADRNALEAVLQAI PAENVLOGVIHAAGALDDGVLDQTTDFSR	4836
QY	66	--APNIT--	70
Db	4837	VLA PKVTGAWNLHETAGNDLAFVFLPSSMSGLLGSAGQSNYAAAANFTLDALAAHRAEG	4896
QY	71	-----ANLTSSL-----	77
Db	4897	LAQSLAWGPMWSGGMAAGLSAALQARLARHGMALGSPAQCTALLGOALARPETQLGAMS	4956
QY	78	-----	77
Db	4957	LDVRAASQASGAAPVPVWRALVRAEARHTAAGAAGALAAARLGAALPEARRADEVKVVQAE	5016
QY	78	-----	77
Db	5017	IARVLSWSAASAVPDRPLSDGLDSLTAVELRNLVLRQVGTLPATLAFDHTVDALTR	5076
QY	78	-----	77
Db	5077	WLLDKVLAVAEPSVSSAKSPQVALDEPIAIIIGICRFGGVADPESFMRLLLEEGSDAVV	5136
QY	78	-----	77
Db	5137	EVPHERNIDAFYDPPDVRGKMTTRFGGFLSDIDRDPAPFGISPREATTMDPQORLLL	5196
QY	78	-----	77
Db	5197	ETSWEAFERAGILPERLMGSDTVGVGLFYQEVAAALAGGIEAFDGYLGTGTTASVASGRI	5256
QY	78	-----	77
Db	5257	SYVLGLKPSLTVDTACSSSLVAVHLACQALRRGECVSALAGGVALMLTPATFVEFSRLR	5316
QY	78	-----	77
Db	5317	GLAPDGRKSFSAADGVHSGCAMLLKPLRDAQRDGDPILAVIRGTAVNQDGRSNGL	5376
QY	78	-----	77
Db	5377	TAPNGSSQOEVIIRALEOAGLADPADVSVYVECHGTGTTGLDPIEVQALGAVLAQGRPSDRP	5436
QY	78	-----	77
Db	5437	LVIGSVKSNIGHTQAAAAGVAGVIKVALALERGLIPRSLHFDAPNPHIPWSELAVQAAKP	5496
QY	78	-----	77
Db	5497	VEMTRNGVPRRAGVSGVGTNAHVVLEEAPAAAFAPAAARSANELFVLSAKSAAALDAQ	5556
QY	78	-----	77
Db	5557	AARLSAHVVAHPELGLDIAFSIATTRS PMTYRLAVAATSREALSALDIAAQQGAPPAA	5616
QY	78	-----LWSC-----GWS-----	84
Db	5617	ARGHASTGSAPKVVFPFGGSQLCMGOKLLSEEPVFRDALSACDRAIQAEAGWSLLAE	5676
QY	85	-----	84
Db	5677	LAADETTSQGRIDVVQPALFAIEVALSALWRSGVPEPDVVVHSGWGEVAAAHAHAGALS	5736
QY	85	-----	84
Db	5737	EDAVAILCRSLLRRISGGEMANVVELSLAEAEALLGYEDRSLSVAVSNSRPRSTVLAGE	5796
QY	85	-----	84
Db	5797	PAALAEVLAIIAAKGVFCRRVKVDVASHSPQIDPLRDELLAALGELEPRQATVMSRSTV	5856
QY	85	-----	84

Db	5857	STIMAGPELVASYWADNRQPVRFAEAVQSLMEDHGFLFVEMSPHPILTTSVTEEIRATK	5916
QY	85	-----	84
Db	5917	REGVAVGSLRRQODERLSMLEALGALWVHGOAVCHWERLFSAGGAGLRVRPLPTYPWQER	5976
QY	85	-----QTNPE-----	90
Db	5977	YVVDAPTGAAGSRFAHAGSHPLLGEMQTLSTQRSTRVWETTLDLKRLPWLGDHRVGOA	6036
QY	91	-----DDTD-----	94
Db	6037	VVFPGAAYLEMALSGAEALGGPLQVSDVVLAELAFADDDTAAVQVMATEERPGRLOF	6096
QY	95	-----PGH-----	97
Db	6097	HVASRVPCHGGNAFRSHARGVLRQIERAEVPARLDLALRARLARQASAPAAATYAALAE	6156
QY	98	-----	97
Db	6157	LEYGPAFQGLVELWRGEALGRVRLPEAAGSPAACRLHPALLDACFHVSSAFADRG	6216
QY	98	-----	97
Db	6217	PWVPVEIGSLRWFORPSGELMCHARSVSHGKPTPDRRSTDFVWVVDSTGAI VAEISGLVAQ	6276
QY	98	-----	97
Db	6277	LAGGVRRREDDWFMPEPAWETA VPGSEVMAGRMLLIGSGGGLGNALHSALT	6336
QY	98	-----AD 99	
Db	6337	HATGRGTSAAQLQALLTASFDGQAPTSVVHLGSLDERGVLDADAPFDADALEESLRGCD	6396
QY	100	LVLY-----ITR-----	106
Db	6397	SVLTVQAVAGAFRDPRLWLVTGQAATGAGDVSAQAAPLLGLGRVIALEHAE	6456
QY	107	-----	106
Db	6457	IDLPARRDGEVDELLAEALLDAEEBVAFRGGERRVARLVRLPETDCREKIEPAEGR	6516
QY	107	FDLEL-----	111
Db	6517	FRLEIDGSGVLDLVLRA TERRPPGGEVEIAVEAAGLNFLDVMRAMGIYPGPGDGPVAL	6576
QY	112	-----	111
Db	6577	GAECGRIVAMGEGVESLRIGQDVAVAPFSGTHVTIDARMLAPRPAALTAQAQNALPV	6636
QY	112	-----	111
Db	6637	AFMTAWYGLVHLGRLRAGERVLIHSATGGTGLAAVQIARHLGAEIFATAGTPEKRAWLRE	6696
QY	112	-----PDGN-----	115
Db	6697	QGI AHVMDSRSLDFAEQVLAATKGEVDVVVLSLSGNAIDASLSTLVDPGRFIELGKTDI	6756
QY	116	-----	115
Db	6757	YADRSGLAHFRKSLSYSAVDLAGLAVRRPVERVAALLAEVVDLLARGALQPLPEIFPLS	6816
QY	116	-----	115
Db	6817	RAADAFRMAQAQOHLKLVLALEDPPVRI RVPGESGVAIRADGAYLVGTGLGGLGSVAG	6876
QY	116	-----ROY-----	118
Db	6877	WLAEOGAGHLVLYGRSAGVSAEQOTAVAALEAHGARVTVARADVADBAQMERILREV	6936
QY	119	-----RGVTQLGG-----	126

Db 6937 GMPLRGVVHAAGILDDGLLMOOTPARFRAVMAKVRGALHLHALTREAPLSFFVLYASGA 6996
 QY 127 -----ACSTPTWCL----- 135
 Db 6997 GULGSPGQNYAANTFLDALAHRRRAOGLFALSIDWGLFADVGLAAGQQNRGARLVTRG 7056
 QY 136 ---ITEDTG---FD----- 143
 Db 7057 TRSLTPDEGLWALERLLDGRDRTQAGVMPFDVQWVEFYFPAASRRRLSLRLMTAPRVASGR 7116
 QY 144 ----- 143
 Db 7117 LAGDRDLLERLATAEAGARAGMLQEVVRAQVSVLRSEGKLDVDAPLTSLGMDSLMGL 7176
 QY 144 -----LGVTI 148
 Db 7177 LRNRIEAVLGITM 7189

RESULT 11

US-09-568-472-5
 ; Sequence 5, Application US/09568472
 ; Patent No. 6358719
 ; GENERAL INFORMATION:
 ; APPLICANT: Schupp, Thomas
 ; APPLICANT: Ligon, James
 ; APPLICANT: Molnar, Istvan
 ; APPLICANT: Zirkie, Ross
 ; APPLICANT: Cyr, Devon
 ; APPLICANT: Goerlach, Joern
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 ; FILE REFERENCE: 4-30582A
 ; CURRENT APPLICATION NUMBER: US/09/568,472
 ; PRIOR FILING DATE: 2000-05-10
 ; PRIOR APPLICATION NUMBER: 09/335,409
 ; PRIOR FILING DATE: 1999-06-17
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 5
 ; LENGTH: 7257
 ; TYPE: PRT
 ; ORGANISM: Sorangium cellulosum
 US-09-568-472-5

Query Match 27.0%; Score 209; DB 4; Length 7257;
 Best Local Similarity 1.5%; Pred. No. 3.1;
 Matches 93; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;
 QY 1 AAGG----- 4
 Db 877 AAGGSVMPGVFPPTAGRRVPLTPYQRYWIEAPAEGLGATTAADALAQNIFYRVDWPEM 936
 QY 5 -----ILH----- 7
 Db 937 PRSSVDSRRARRSGWLVLADRGVGEAAAAALSGGSCAVLHAPAEASVAEQVTOALG 996
 QY 8 ----- 7
 Db 997 GRNDQGVLYLWGLDVAVEAGASAEVAKVTHLAAAPVLAIQALGTGPRSLWIVTRG 1056
 QY 8 ----- 7
 Db 1057 ACTVGEPPDAAPCOAALWGMGRVALEHPGSGWGLVDLDPEESPTEVEALLVALLSPDAE 1116
 QY 8 ----- 7
 Db 1117 DOLAFROGRRRAARLVAAPPEGNAAPVLSAEGSYLVGTGGLGALLVARLWVERGAGHL 1176
 QY 8 -----LE-----LLVAVG 15
 Db 1177 VLISRHGLPDREKNGRDPPEVRARIAIEALEAQGARVTVAADVADAEAGMAALLAAVE 1236
 QY 16 P-----DVFOAHO----- 23

Db 1237 PPLRGVVAAGLLDDGLLAHODAGRLARVLRPKVEGAWVLHTLTREQPLDLFVLPSSASG 1296
 QY 24 ----- 23
 Db 1297 VFGSICQGSYAAGNAPFLDALADLRRTOGLAALSIAWGLWAEGBGMSQORREHSGIWA 1356
 QY 24 ----- 23
 Db 1357 MPTSRALAAMEWLLGTRATORVVIQMDWAHAGAAPRDSRGRFMDRLVTATKEASSSASP 1416
 QY 24 ----- 23
 Db 1417 AVERWRNASVETRYSALVELYRGVYAGVMGFTDQGLDVRGFAEOGLSLMAVEIRKRL 1476
 QY 24 ----- 23
 Db 1477 QGELGMLSATLAFDHPHTVERLYLLSQALEIQDRTDVRSVRLPATEDDPIAIVGAACRF 1536
 QY 24 ---EDTERY----- 29
 Db 1537 PGVEDLESYWQLLTEGVVYSTVTPADRMNGADGRVPGSGEAOQTQTVVPRGGFLREVET 1596
 QY 30 ----- 29
 Db 1597 DAFFPHISPREAMSLDPQORLLLEVSWEAIERAGODPSALRESPTGTVFGAGPNEYAERV 1656
 QY 30 ----- 29
 Db 1657 QELADEAAGLYSGTGNMLSVAAGRLSFFLGLHCPPLAVDTACSSSLVALHLGCOSLRRE 1716
 QY 30 ----- 29
 Db 1717 CDQALVGGVNMLLSPKTFALLSRMHALSPGGRCKTFADADGYARAEGCAVVVVKRLSDA 1776
 QY 30 ----- 29
 Db 1777 QDRDPIAVIRGTAINHDPSSGLTVPSGPAQEALLRQALAHAGVVPADVDFVECHGTG 1836
 QY 30 ----- 29
 Db 1837 TALGDPFIEVRALSDVYGOARPADRPLILGAANKANLGHMEPAAGLAGLLKAVLALGOEIP 1896
 QY 30 ----- 29
 Db 1897 AQPELGELNPLLPWEALPVAVARAAPVWPRTDRPRFAGVSSFGMSGTNAHVLEAPAVE 1956
 QY 30 ----- 29
 Db 1957 LWPAAPERSAELLVLSGSEKALDAQARLREHLDHMPHGLGDLVAFSLAGTTRSAMNHRL 2016
 QY 30 ----- 29
 Db 2017 AVAVTSREGLAALSAVAQGTTPGAARCIASSSRGKLAFLTGQGAOTPGMGRGLCAAW 2076
 QY 30 ----- 29
 Db 2077 PAFREAFDRCVALFDRELDRLREVWMAEPGSAESLLLDOTAFQPALFTVEYALTALWR 2136
 QY 30 ----- 29
 Db 2137 SWGVEPELVAGHSAGELVAACVAGVPSLEQVRLVAARGRLMOGLSAGGANVSLGAPEAE 2196
 QY 30 ----- 29
 Db 2197 VAAAVAPHAASVSTAANGPEQVVIAGVEQVAVOIAAGFAARGARTKRLHVSHASHSPLM 2256
 QY 30 -----VLTNLN----- 35
 Db 2257 EPMLEFRVAASVTYRPPSVSLVSNLSKGVVADELSAPGYWVRHVREAVRFADGVKALH 2316
 QY 36 ----- 35

Db 2317 EAGAGTFVEVGPKPTLLGILLPACLPENAEPTLLASLRAGREAAAGVLEALGRMAAGGSVS 2376
QY 36 ----- 35
Db 2377 WFCVFTAGRRVPLPTYPHQRYWPDIEPDSRRHAAADPTQGWFYRDWPEIPRSLOKS 2436
QY 36 ----- 35
Db 2437 BEASRGSWLVLADKSGVGEVAAALSTRGLPCVVLHAPAEYSATAELVTEAAGGRSDQV 2496
QY 36 ----- 39
Db 2497 VLYLWGLDAVGAESIDEIGDATRRATAPVLGLARFLSTVSCSPRLWVTRGACIVGDE 2556
QY 40 ----- 39
Db 2557 PAIAPCOAALMGMRVAALHPGAMGGLVDLDRASPQASPIDGEMLVTELLSOETEDQ 2616
QY 40 ----- 39
Db 2617 LAFRHRRHARLVAAPQGOAAPVLSAEASYLVGTGLGGLIIVAQWLVELGARHLVL 2676
QY 40 ----- 39
Db 2677 TSRRGLPDRQWCEQPPEIRARIAAEALEARGARTVVAADVADVEPMTALVSSVEPP 2736
QY 40 ----- 39
Db 2737 LRGVVHAAGVSMRPLAETDETLLESVLPRPKVAGSWLLHRLHGRPLDLFLVSSGAAMV 2796
QY 40 ----- 39
Db 2797 GSHSOGAAANAFLDGLAHLRRSQLPALSVAWGLWAEGMADAAHARLSDIGVLPMS 2856
QY 40 ----- 39
Db 2857 TSAALSALQRLVETGAAQRTVTRMDWARFAPVYTARGRNLLSALVAGRDIIAPSPAAA 2916
QY 40 ----- 39
Db 2917 TRNWRGLSVAEARVALHEIVHGAVARVLGFLDPSALDPGMGFNEQGLDSLMAVEIRNLQ 2976
QY 40 ----- 39
Db 2977 AELDVRLSTTLAFDHTVQRLVHELLVDVLKLEDRSDTQHVRSLASDEPIAIVGAACRFP 3036
QY 40 ----- 39
Db 3037 GGVEDLESYQWLLAEGVVVSAEVPADRWDADWDYDPEIPGRTYVTKGAFRLDLQLDA 3096
QY 40 ----- 39
Db 3097 TFRISPREAMSLDPOORLLLESWEALESAGIAPDTRDSPTGVFVGAGPNEYTQRLR 3156
QY 40 ----- 39
Db 3157 GFTDGAAGLYGGTGNNMLSVTAGRLSPFLGLHGPTLAMDTACSSSLVALHLCQSLRLGEC 3216
QY 40 ----- 39
Db 3217 DOALVGVVNLAPETVLLSRMRALSPOGRCKTFSADADYARGEGCAVVVLKRLRDAQ 3276
QY 40 ----- 42
Db 3277 RAGDSLALIRGSVNHDPSSGLTVPNGPAQOALLRQALSQAGVSPVDVDFVECHGTGT 3336
QY 43 ---DP--- 44
Db 3337 ALGDPIEQVQALSEVYGFGRSGDRPLVLGAANKANVAHLEAASGLASLLKAVLALRHEQIPA 3396
QY 45 ----- 44
Db 3397 QPELGINPHLPWNTLPLVAVPRKAVPWGRGARPRRAGVSAFGLSGTNVHVLEEAPEVEP 3456

QY 45 ----- 44
Db 3457 APAAPARPVELVVLSAKSAALDAAAARLSAHLSAHPSELGDAVAFSLATTRSPMEHRLA 3516
QY 45 ----- 44
Db 3517 IATTSREALRGALDAAAOQKTPQGAVRGKAVSSRGKLAFLFTGOGAQMPCMGRLYETWP 3576
QY 45 ----- 44
Db 3577 AFREAPDRCVALLFDREIDOPUREVWMAAPGLAQAAARLDOTAYAPALFALEYAALWRS 3636
QY 45 ----- 44
Db 3637 MGVEPHVLLGHSIGELVAACVAGVPSLEDAVRLVAARGRLMQALPAGGANVAIAASEAEV 3696
QY 45 ----- 50
Db 3697 AASVAPHAATVSIAAVNGPDVAVIAGAEVQVLALCATPAARGIRTKRLAVSHAFHPLMD 3756
QY 51 ----- 50
Db 3757 PMLEDFORVAATIAYRAPDRPVSVNVTHGVAGPEIATPEYVVRHVRSVAVRFGDGAKALHA 3816
QY 51 ----- 50
Db 3817 AGAATFVEVGPKVLLGLLPACLGADAVLVPSURADRSECEVVLAALGAWYANGGALDW 3876
QY 51 ----- 52
Db 3877 KGVFPDGARRVLMYPMQORERHMDLTPRSAAPAGIAGRWPLAGVGLCMPGAVLHHVLS 3936
QY 53 ----- 55
Db 3937 IGPRHQPFLGDLHVGKVVVPGAVHVAVILSIAERWPERAIBLTGVFEFLKATAMEPDQE 3996
QY 56 ----- 55
Db 3997 VELHAVLTPTEAGDGYLFELATLAAPETERRTTHARGVQPTDGAPGALPRLEVLDR 4056
QY 56 ----- 55
Db 4057 IQPLDFAGFLDRLSVAIRIGHGPLWRWLQDGRVGDSEASLATLVPTYNAHDVAPLHPILLD 4116
QY 56 ----- 55
Db 4117 NGFAVSLSTRSEPEDDGTPLPFAVERVRWRAPVGRVRCGGVPRSQAGVSSFVLVDE 4176
QY 56 ----- 55
Db 4177 TGEVVAEVEGFVCCRAPPREVFLROESGASTAALYRLDWPPEALPDAPAERIEESWVVAA 4236
QY 56 ----- 55
Db 4237 PGSEMAAALATRLNRCVLAEPKLEAALAGVSPAGVICLWEAGAHEEAPAAQORVATEGL 4296
QY 56 ----- 55
Db 4297 SVVQALDRDRAVRLMWVTMGAVAVEAGERVOVATAPVWGLGRTVMOERPELSCTLVLDPE 4356
QY 56 ----- 65
Db 4357 ADAARSADVLLRELGRADDETQVAFRSGKRRVARLVKATTPGELLVPPDAESYRLEAGQKG 4416
QY 66 ----- 65
Db 4417 TLDQLRAPAQRAPGEGVEIKVTASGLNFRFTVLAVLGMYPGDAGPMGDCAGVATAVG 4476
QY 66 ----- 65
Db 4477 QGVRHVAVGDVNMVTLGTLHRFTVTDARLVVROPAGLTPAQAAATVPVAFLTAWLALHDLGN 4536

```
QY 66 ----- 65
Db 4537 LRRGERVLIHAAAGVGMAAVQIARWIGAEVFATASPSKWAAVQAMGVPRTHIASRTLE 4596
QY 66 ----- 65
Db 4597 PAETFRVOTGGRGVVVLNALAGEFVDASLSLLSTGGRFLEMGTDIRDRAAVAAAHPGV 4656
QY 66 ----- 65
Db 4657 RYRVFDILELAPDRTREILERVEGFAAGHLRALPVHAFAITKAEAAFRMAQARHOGKV 4716
QY 66 ----- 65
Db 4717 VLLPAPSAAPLAPGTGTVLLTGGLGALGHVARMLAQGVPHMYLTGRRGLDTPGAAKAVA 4776
QY 66 ----- 65
Db 4777 EIEALGARVTIAASVDADRNALEAVLQAI PAEMPLOGVIAHAGALDDGVLEQTTDRFSR 4836
QY 66 --APNIT----- 70
Db 4837 VLAPKVTGANWLHETAGNDLAFVFLFSSMSGLLGSAGOSNYAANTFLDALAAHRAEG 4896
QY 71 -----ANLTSSL----- 77
Db 4897 LAAQSLAWGPFSDCGMAAGLSAALQARLARHGMGALS PAQTALLGQALARPETQLGAMS 4956
QY 78 ----- 77
Db 4957 LDVRAASQASGAAPPVWRALVRAEARHTAAGAQAAGALAAARLGALEARRADEVRKVQAE 5016
QY 78 ----- 77
Db 5017 IARVLSAASAVPDRPLSLDLGLSLTAVELRNVLGQVGTATLPATLAFDHPPTVDALTR 5076
QY 78 ----- 77
Db 5077 WLLDKVLVAEPSVSAKSSPOVALDEPIAIIIGICRFPFGVADPESFWRLEEGSDAVV 5136
QY 78 ----- 77
Db 5137 EVPHERWDIDAFYDPPDVRGKMTTRFGGFLSDIDRDPAPFGISPREATTMDPQORLLL 5196
QY 78 ----- 77
Db 5197 ETSWEAFERAGILPERLMGSDTGVFVGLFYOEYALAGIEAFDYLGTGTTASVSGRI 5256
QY 78 ----- 77
Db 5257 SYVLGLKGPSLTVDTACSSSLVAVHLACOALRRGECVALAGGVALMLTPATFVEFSRLR 5316
QY 78 ----- 77
Db 5317 GLAPDRCKSFAAADGVGWSGCCAMLLKPLRDAQRDGPILAVIRGTAVNQGRSNGL 5376
QY 78 ----- 77
Db 5377 TAPNGSSQOEVIIRALEOAGLAPADVSVECHGTGTTLGDPIEQVALCAVLAQGRPSDRP 5436
QY 78 ----- 77
Db 5437 LVIGSVKSNIGHTQAAAGVAGVIKVALALERGLIPRSLHFDAPNPHIPWSELAVQVAAKP 5496
QY 78 ----- 77
Db 5497 VENTRNGVPRAGVSSFGVSGTNAHVLEEAFAAFAAARSALFVLSAKSAALDAQ 5556
QY 78 ----- 77
Db 5557 AARLSAHVVAHPELGLDLAFSLATTRS PMTYRLA VAATSREALSAAALDTAAQCPAA 5616
QY 78 -----LSVC-----GWS----- 84

Db 5617 ARGHASTGSAPKVVVVFPPGOGSOMLGMQKLLSEEPVFRDALSACDRAIOAEGMSLLAE 5676
QY 85 ----- 84
Db 5677 LAADETTSQLGRIDVQVQALPAIEVALSALWRSWGVPEPDAVVGHSMGEVAHAHVAGALS 5736
QY 85 ----- 84
Db 5737 EDVAIIICRRSLRLRRISGOGEMAVVELSLAEAEALLGYEDRLSVAVSNSPRSTVLAGE 5796
QY 85 ----- 84
Db 5797 PAALAEVLALIAAKGVFCRRVKVDVASHSPOIDPLRDELLAALGELEPROATVSMRSTVT 5856
QY 85 ----- 84
Db 5857 STIMAGPELVASYWADNVROPVRAEAVOSLMEDGHCLFVEMSPHPLITTSVEIRRA TK 5916
QY 85 ----- 84
Db 5917 REGVAVGSLRRGODERLSMLEALGALWVHGQAVGWERLFSAGGAGLRRVPLTPYQORER 5976
QY 85 -----OTINPE----- 90
Db 5977 YWVDAPTGGAAGSRFAHAGSHPLLGEMOTLSTQSTRVWETTLDKRLPWLGDHRVOGA 6036
QY 91 -----DDTD----- 94
Db 6037 VVFPGAAYLEMAUSSGAELGDGPLOVSDVLAELAFADDTTPAAVQVWATBERPORLOF 6096
QY 95 -----PGH----- 97
Db 6097 HVASRPVGHGGAARFRSHARGVLRQIERAEVPARLDIAALARLQASAPAAATYAALAE MG 6156
QY 98 ----- 97
Db 6157 LEYPAFOGLVELWRGEALGRVLRPEAGSPAACRLHPALLDACFHVSSAFADRGAT 6216
QY 98 ----- 97
Db 6217 PMVPEIGSLRWFORPSGELMCHARSVSHGKTPDRRSTDFWVVDSTGAIVAEISGLVAQ 6276
QY 98 ----- 97
Db 6277 RLAGGVRRREDDWMEPAWEPTAVPGSEVMAGRMLLIGSGGLGAALHSALTAEAGHSV 6336
QY 98 -----AD 99
Db 6337 HATGRGTSAAGLQALLTASFQQAPTS VVHLGSLDERGVLDADAFDADALPESLVRGCD 6396
QY 100 LVLY-----ITR----- 106
Db 6397 SVLMTVQAVAGAGFRDPPRLMLVTRGAOIGA GDSVVAQAPLLGLGLRVIALEHAELRCAR 6456
QY 107 ----- 106
Db 6457 IDLDPARRDEVDLELLAELLADDAEEVAFRGGERRVARLVRRLPETDCREKIEPAERGP 6516
QY 107 FOLEL----- 111
Db 6517 FRLEIDGSGVLDLVL RATERPPGPEVEIAVEAAGLNFUDVMRAMGIYPGPGDGPVAL 6576
QY 112 ----- 111
Db 6577 GAECSGRIVAMGEVGSURIGQDVAVAPFSFGTHVTIDARMLAPRPAALTAQAALLPV 6636
QY 112 ----- 111
Db 6637 AFMTAWYGLVHLGRLRAGERVLIHSATGGTGLAAVQIARHLGAEI FATAGTPEKRAWLRE 6696
QY 112 -----PDGN----- 115
```

Db 6697 OGIAHVWDSRLDFAEQVLAATKGEQVDVVLNSLSGAAIDASLSTLVPDGRFIELGKTDI 6756
Qy 116 ----- 115
Db 6757 YADRSGLAHFRKLSYSADVLDLAGLAVRRPERVAALLAEVVDLLARGALQPLPVEIPPLS 6816
Qy 116 ----- 115
Db 6817 RAADAFRMAQAQHLGKLVLAELEDPVIRVPGESGVAIRADGAYLVTGGLGGLSLVAG 6876
Qy 116 ----- 118
Db 6877 WLAEOGAGHLVLVGRGAVSAEQQTAVAALAEHAGARVTVARADVADRAQMERILREVNTAS 6936
Qy 119 ----- 126
Db 6937 GMPLRGVVHAAGILDDGLLMQOTPARFRAVNAKPVKGALHLHALTREAPLSFFVLYASGA 6996
Qy 127 ----- 135
Db 6997 GLLGSPGOGNYAANTFLDALAHRRRAOGLPALSIDMGLFADVGLAAGQONRGARLVTRG 7056
Qy 136 ----- 143
Db 7057 TRSLTPDEGLWALERLLDGRTOAGVMPFDVQWVEFYPAASRRRLSRLMTARRVASGR 7116
Qy 144 ----- 143
Db 7117 LAGDRDLERLATAEAGARAGMLOEVVRAQVSVLRLSEGKLDVDAPLTSIGMDSLMGLE 7176
Qy 144 ----- 148
Db 7177 LRNRIEAVLGITM 7189

RESULT 12

US-09-567-899-5
; Sequence 5, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-899-5

Query Match 27.0%; Score 209; DB 4; Length 7257;
Best Local Similarity 1.5%; Pred. No. 3.1;
Matches 93; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;
Qy 1 AAGG----- 4
|||
Db 877 AAGGSVWPGVFTAGRVPLPTYPMQORQRYWIEAPAEGLGATAADALAQMRYRVDWPEM 936
Qy 5 ----- 7
||| ILH-----
Db 937 PRSSVDSRRARGMLVLDRGVGEAAAAALSSQCCSCLVLAHAESAIAVEQVTOALG 996
Qy 8 ----- 7

Db 997 GRNDWQGVLYLNGLDRAVVEAGASABEVAKVTHLAAAPVLALQALGTQPRSPRLWIVTRG 1056
Qy 8 ----- 7
Db 1057 ACTVGGEPPDAA PCQAALWCMGRVAALHHPGSGGLVDLPBESPTVEVALVAELLSPOAE 1116
Qy 8 ----- 7
Db 1117 DQLAFQGRRRRAARLVAAPPEGNAAPVSLSAEGSYLVTGGLGALGLLVARMLVERGAGHL 1176
Qy 8 ----- 15
||| LLVAVG
Db 1177 VLISRHGLPDREWGRDQPPPEVRARIAAEIAEAQAGARTVAADVADAEGMAALLAAVE 1236
Qy 16 P-----DVFOAHQ----- 23
|||
Db 1237 PPLRGVHAAGLLDDGLLAHQDAGRLARVLRPKVEGAWVLHTLTREQPLDLFLVLFSSASG 1296
Qy 24 ----- 23
Db 1297 VFGSIGGSYAAGNAFLDALADLRLTQGLAALSTAWGLWAEGGMSQORREHEASGTWA 1356
Qy 24 ----- 23
Db 1357 MPTSRALAAMEWLLGTRATORVVIQMDWAHAGAAPROASRGRFWDRLVTATKEASSAVP 1416
Qy 24 ----- 23
Db 1417 AVERWRNASVETRSALYELVRGVVAGVMGFTDQGLDVRGRFAEOGLSLMAVEIRKRL 1476
Qy 24 ----- 23
Db 1477 QGELGMPLSATLAFDHTPTVERLVEYLLSQALELODRTDVRSVRLPATEDPIAIVGAACRF 1536
Qy 24 ----- 29
||| EDTERY
Db 1537 PGVEDLESYMWQLLTEGVVYSTEVPADRMNGADGRVPGSGEAOQTVPVPRGGLREVET 1596
Qy 30 ----- 29
Db 1597 DAFFHISPREAMSLDPOORLLLEVSWEATERAGQDPSALRESPTGVFGVAGPNEYAERV 1656
Qy 30 ----- 29
Db 1657 QELADEAAGLYSGTGNMLSVAAGRLSFFLGLHGPTLAVDTACSSSLVALHLGCQSURRGE 1716
Qy 30 ----- 29
Db 1717 CDQALVGGVNMILLSPKTFALLSRMHALSPEGRCCKTFSADADGYARAEGCAMVVLKRLSDA 1776
Qy 30 ----- 29
Db 1777 QRDRDPILAVIRGTAINHDGSSGLTVPSGPAQEALLRQALAHAGVVPADVDVECHGTG 1836
Qy 30 ----- 29
Db 1837 TALGDPIEVRLSDVYGOARPADRPILGAAKANLGHMEPAAGLAGLLKAVLALGOEQIP 1896
Qy 30 ----- 29
Db 1897 AQPGLGELNPLLPWEALPVAVARAAPVWPRTDRPRFAGVSGVSGMGTNAHVVEEAPAVE 1956
Qy 30 ----- 29
Db 1957 LWPAAPERSAELLVLISGKSEGALDAQARLREHLDHMPGLGDVAFSLATTRSANNHRL 2016
Qy 30 ----- 29
Db 2017 AVAVTSREGLLAALSAVAQOQTTPGAARCIASSSRGKLAFLFTGOGAQTGPMGRGLCAAW 2076
Qy 30 ----- 29

```
Db 2077 PAFREAPDRCVALLFDRELDRLREVWMAEPGSAESLLLDQTAFTQPAFTVEYALTALWR 2136
QY 30 ----- 29
Db 2137 SWGVEPELVAGHSAGELVAACVAGVFSLEDGVLVAARGRLMOGLSAGGAMVSLGAPEAE 2196
QY 30 ----- 29
Db 2197 VAAAVAPHAASVIAAVNGPEQVVIAGVEQAVQAIAGFAARGARTKRLHVSASHSPLM 2256
QY 30 ----- VLTNLN ----- 35
Db 2257 EPMLEBFRVAASVTYRRFSVLSVNLGKWADELSAPGYVVRHVREAVRFADGVKALH 2316
QY 36 ----- 35
Db 2317 EAGAGTFVEVGPKTLLGLLPACLPPEAEPTLLASLPAGREEAAGVLEALGRLLWAAGGSVS 2376
QY 36 ----- 35
Db 2377 WPCVPTAGRRVLPPTYPMQORQYWPDIPEPSRRHAAADPTQGWFYRVDWPEIPRSLQKS 2436
QY 36 ----- 35
Db 2437 EBASRGSWLVLADKGVGGEVAAALSTRGLPCVVLHAPAEATSATAELVTEAAGGRSDWQV 2496
QY 36 ----- IGAE ----- 39
Db 2497 VLYLWGLDAVWGAESAIDEIGDTRRATAPVLGLARFLSTVSCSPRLWVTRGACIVGDE 2556
QY 40 ----- 39
Db 2557 PAIAPCOAALWGMGRVAALHPGAWGGLVDLPDPRASPQASPIDGEMLVTELLSQETEDQ 2616
QY 40 ----- 39
Db 2617 LAFRHRRRHAARLVAAPQGOQAAPVLSAEASLYVTGGLGLLIVAOWLVELGARHLVL 2676
QY 40 ----- 39
Db 2677 TSRRGLPDROAWCEQPPEIRARIAAVEALEARGARTVAAVDVADVEPMTALVSSVEPP 2736
QY 40 ----- 39
Db 2737 LRGVVHAAGVSMRPLAETDETLLLESVLRPKVAGSWLLHRLHGRPLDLFVLFSSGAAV 2796
QY 40 ----- 39
Db 2797 GSHSQGAVAAANAFDGLAHLRRSOSLPALSVANGLWAEGMADAHAHRLSDIGVLPM 2856
QY 40 ----- 39
Db 2857 TSAALSALQRLVETGAAQRTVTRMDWARFAPVYTARGRRNLLSALVAGRDIIAPSPAAA 2916
QY 40 ----- 39
Db 2917 TRNRGLSVAEARVALHEIVHGAVARVLGFLDPSALDPGMEQGLDSLMAVEIRNLQ 2976
QY 40 ----- 39
Db 2977 AELDVRLSTTLAFDHTVQRLVEHLLVDVLKLEDRSDTQHVRSIASDEPTAIVGAACRFP 3036
QY 40 ----- 39
Db 3037 GGVEDLESYQOLLAEVWVSNEVPADRWDAAWDYDPDPEIPGRYTVTKGAFRLDRLQDA 3096
QY 40 ----- 39
Db 3097 TFFRISPREAMSLDPQORLLLEVSWEALESAGIAPDLRDSPTGVFVGAGPNEYYYQRLR 3156
QY 40 ----- 39
Db 3157 GFTDGAAGLYCGTGNMLSVTAGRLSFFLGLHGLPTLMDTACSSSLVALHLACOSLRIGEC 3216
QY 40 ----- 39
Db 3217 DOALVGVNVLLAPETTFVLLSRMALSPDGRCKTFSDADGARGEGCAVVVLKRLRDAO 3276
QY 40 ----- LLR ----- 42
Db 3277 RAGDSIILALIRGSVNHDPGSSGLTVPNGPAQALLRLQALSOAGVSPVDVFVECHGTGT 3336
QY 43 ----- DP ----- 44
Db 3337 ALGDPIEVQALSEVYGPGRSDRPLVLGAANKANVAHLEAASGLASULLKAVLALRHEQIPA 3396
QY 45 ----- 44
Db 3397 QPELGEINPHLPWNTLPLVAVPRKAVPMWGRGARPRRAGVSFGLSGTNVHVLEAPEVEP 3456
QY 45 ----- 44
Db 3457 APAAPRPVELVVLSSAKSAAALDAAAARLSAHLSAHPELSLGDVAFSLATTRSPMEHRLA 3516
QY 45 ----- 44
Db 3517 IATTSREALRGALDAAAQOKTPOGAVRGKAVSSRGKLAFLFTGQGAQPMGMRGLYETWP 3576
QY 45 ----- 44
Db 3577 AFREAFDRCVALFDREIDQPLREVWMAAPGLAQARLDOTAYAQPALFALEYALALWRS 3636
QY 45 ----- 44
Db 3637 WGVPHVLLGHSTIGELVAACVAGVFSLEDVAVRLVAARGRLMOALPAGGANVAIAAEAEV 3696
QY 45 ----- SLGAQF ----- 50
Db 3697 AASVAPHAATVIAAVNGPDVAVIAGAEVQVLALGALTFAARGIRTKRLAVSHAFHPLMD 3756
QY 51 ----- 50
Db 3757 PMLEDQFQVAATIAIRAPDRPVVSNVTGHVAGPEIATPEYVVRHVRSVAVRFGGAKALHA 3816
QY 51 ----- 50
Db 3817 AGAATFVEYGPXPKVLLGLLPACLGEADAVLPSLRADRSCEVVLALGAMVANGCALDW 3876
QY 51 ----- RV ----- 52
Db 3877 KGVFPDGAARRVALPMYPMQREHMDLTFRSAAPAGIAGRWPAGVGLCMPCGVLHHVLS 3936
QY 53 ----- HLV ----- 55
Db 3937 IGPRHQPFGLDHLVFGKVVVPGAFHVAVILSIAAERWPERAIELTGVEFLKAIAMEPDOE 3996
QY 56 ----- 55
Db 3997 VELHAVLTPEAAGDGLFELATLAAPETERRWTTHARGVQPTDGAPCALPRLEVLEORA 4056
QY 56 ----- 55
Db 4057 IQPLDFAGFLORLSAVRIGWGLWRWLODGRVGDSEASLATLVPTYNNAHDVAPLHPILLD 4116
QY 56 ----- 55
Db 4117 NGFAVSLLSRSEPDGTPPLPFAVERVWRWRAPVGRVCGGVPRSQAFVCSFVLVDE 4176
QY 56 ----- 55
Db 4177 TCEVVAEVEGVFCRRAPREVFLROESGASTAALYRLDWPAPLPDAPAEITEESWVVA 4236
QY 56 ----- 55
Db 4237 PGSEMAAALATRLNRCVLAEPKGLEAALAGVSPAGVITCLWEAGAHEEPAAAOORVATEGL 4296
```

```
QY 56 ----- 55
Db 4297 SVVQALRDRVRLMWVWTCGAVAVAGERVQVATAPWGLGRVTVMQERPELSCITLVLEPE 4356
QY 56 ----- 65
Db 4357 ADAARSADVLLRELGRADDETQVAFRSGKRRVARLVKATTPEGLLVPDAESYRLEAGQKG 4416
QY 66 ----- 65
Db 4417 TLDQLRLAPQORRAPGPEVEIKVTASGLNFRVLAFLVGMYPDAGPMGGDCAGVATAVG 4476
QY 66 ----- 65
Db 4477 QGVRHVAVGDAVMTGLTLHRFVTVUDARLVVROPAGLTTPAQAAATVPVAFLTAWLALHDLGN 4536
QY 66 ----- 65
Db 4537 LRRGERVLIHAAAGVGMAAVQIARWIGAEVFATASPSKMAAVQAMGVPRTHIASRTLE 4596
QY 66 ----- 65
Db 4597 FAETFRQVTGGRGVDVNLALAGEFVDASLSLLSTGGRFLEMGTDIRDRAAVAAAHPGV 4656
QY 66 ----- 65
Db 4657 RYRVFILLELAPDRTEILERVEGFAAGHLRALPVHAPAITKAEAAFRFMAQARHQKV 4716
QY 66 ----- 65
Db 4717 VLLPAPSAAPLAPTGVLLTGGGALGLHVARLWLAQQGVPHVLTGRRGLDTPGAAKAVA 4776
QY 66 ----- 65
Db 4777 ETEALGARVTIAASVDADRNALEAVLQAIPEAWPLQGVTHAAGALDDGVLDQTTDRFSR 4836
QY 66 ----- 70
Db 4837 VLAPKVTGANLHELTAGNDLAFVLFSSMSGLLSAGOSNYAAANTFLDALAAHRAEG 4896
QY 71 ----- 77
Db 4897 LAAQSLANGPWSGGGWAAGLSAALQARLARHGMGALSQAQGTALLGOALARPETQLGAMS 4956
QY 78 ----- 77
Db 4957 LDVRAASQASGAAPVPMRALVRAEARHTAAGACQALAAALGALPEARRADEVRKVQAE 5016
QY 78 ----- 77
Db 5017 IARVLSWSAASAVPVDPRPLSDGLDLSLTAVELRNVLGQRVGATLPATLAFDHPHTVDALTR 5076
QY 78 ----- 77
Db 5077 WLLEDKVLVAEPSVSSAKSSPOVALDEPIAIGICRFFGGVADPESFWRLLLEGSDAV 5136
QY 78 ----- 77
Db 5137 EVPHERWDIDAFYDPDPDVRGKMTTRFGGFLSDIDRFDPAFFGISPREATTMDPOORLLL 5196
QY 78 ----- 77
Db 5197 ETSWEAFERAGILPERLMGSDTGVFVGLFYQEYAAALAGGIEAPDGLGTGTASVAGRI 5256
QY 78 ----- 77
Db 5257 SYVLGLKGPLSLTVDTACSSSLVAVHLACQALRGECSVALAGGVALMLTPATFVFSRLR 5316
QY 78 ----- 77
Db 5317 GLAPDGRCKSFAAADGVGWSGECAMLLKPLRDAORDGDPILAVIRGTAVNQDGRSNGL 5376
QY 78 ----- 77

Db 5377 TAPNSSQQEVIRRALEQAGLAPADVSVECHGTGTTLGDPIEQALGAVLAOGRPSDRP 5436
QY 78 ----- 77
Db 5437 LVIGSVKSNIGHTOAAAGVAGVIKVALALERGLIPRSLHFDAPNPHIPWSELAVOVAKP 5496
QY 78 ----- 77
Db 5497 VENTRNGVPRAGVSFGVSGTNAHVVEEAPAAAAPAAAARSALFVLSAKSAAALDAQ 5556
QY 78 ----- 77
Db 5557 AARLSAHVVAHPGLGLDLAFSLATTRSPMTYRLAVAATSRALSAALDTAAOQOAPPA 5616
QY 78 ----- 84
Db 5617 ARGHASTGSAKPVVFPFGQSQWLMGQKLLSEEPVFRDALSACDRAIQAEAGWSLLAE 5676
QY 85 ----- 84
Db 5677 LADETTSQLGRIDVVQPALFAIEVALSALWRSWGVPEPDAVVGHSMGEVAAAHHVAGALS 5736
QY 85 ----- 84
Db 5737 EDVAIIICRRSLLLRISGOGEMAVVELSLAEAEALLGYEDRLSVAVNSPRSTVLAGE 5796
QY 85 ----- 84
Db 5797 PAALAEVLALIAAKGVFCRRVKVDVASHPSPOIDPLRDELLAALGELEPROQATVMSRSTV 5856
QY 85 ----- 84
Db 5857 STIMAGPELVASYADNVQPVRAEAVQSLMEDGHGLFVEMSPHILTTTSVEEIRRTK 5916
QY 85 ----- 84
Db 5917 REGVAVGSLRRGQDERLSMLEALCALVWVQAVGWERLFSAGAGLRRVPLTYPWORER 5976
QY 85 ----- 90
Db 5977 YWVDAPTGAAGSRFAHAGSHPLLGEMOTLSTORSTRWETTLDLKLPLWGLDHRVOGA 6036
QY 91 ----- 94
Db 6037 VVFPQAYLENALSSGAELGDPLOVSDVVLAELAFADDTPAAVQVMATEERPGLQF 6096
QY 95 ----- 97
Db 6097 HVASRVPCHGGAAFRSHARGVLQIERAEVPARLDLAALRARLOASAPAAATYAALAE 6156
QY 98 ----- 97
Db 6157 LEYGPFOGLVELWRGEALGRVRLPEAAGSPAACRUHPALLDACFHVSSAFADRGAT 6216
QY 98 ----- 97
Db 6217 PWPVETGSLRWFORPSGELWCHARSVSHGKPTDPRRSTDFWVVDSTCAI VAEISGLVAQ 6276
QY 98 ----- 97
Db 6277 RLAGVRRREDDWFMPEAPEPTAVPGSEVMAGRMLLIGSGGLGAALHLSALTEAGHSV 6336
QY 98 -----AD 99
Db 6337 HATGRGTSAGLQALLTASFDGQAPTSVVHIGSLDERGVLDADAFDADALEESLVRGCD 6396
QY 100 LVLY----- 106
Db 6397 SVLWTVQAVAGAFRDPPLRLWLVTRGCAQIAGAGDVSAQAPLGLGLRVIALEHAEI 6456
QY 107 ----- 106
```


Db 6457 IDLPARRDGEYDELLAELLADDAEEVAFRGERRVARLVRLPETDCREKIEPAEGRP 6516
Qy 107 FDLEL-----111
Db 6517 FRLEIDGSGVLDLVLRTERRPPGGEIATVEAAGLNFLDMRAMGIYPGPGDGPVAL 6576
Qy 112-----111
Db 6577 GAECGRIVANGEVESLRIGQDVVAVAPFSFGTHVTIDARMLAPRPAALTAQAALPV 6636
Qy 112-----111
Db 6637 AFMTAWGLVHLGRLAGERVLIHSATGGTGLAAVOIARHLGAEIFATAGTPEKRWLRE 6696
Qy 112-----115
Db 6697 QGIAHVWDSRLDFAEQVLAATKGEVDVVLNSGAAIDASLTLPDGRFIELGKTDI 6756
Qy 116-----115
Db 6757 YADRSGLAHFRKSLSYSVDLAGLAVRRPERRVAALLAEVVDLLARGALQPLPVEIFPLS 6816
Qy 116-----115
Db 6817 RAADAFRMAQAQHLGLVLALEDPDVIRVPGESVAIRADGAYLVVTGGGLGLSVAG 6876
Qy 116-----118
Db 6877 WLAEOGAGHLVLCRSAGVSAEQQTAVAALEAHGARVTVARADVADRAQMERILREVTAS 6936
Qy 119-----126
Db 6937 GMPLRGVVAAGILDDGLLMQOTPARFRAVMKPVKRGALHLHALTREAPLSFFVLVASGA 6996
Qy 127-----135
Db 6997 GLLGSPGQNYAAANTFLDALHRRRAQGLPALSIDMGLFADVGLAAGQQRGARLVTRG 7056
Qy 136-----143
Db 7057 TRSLTPDEGLWALERLLDGDRTQAGVMPFDVROWVEFYPAASRRRLSRLMTARRVASGR 7116
Qy 144-----143
Db 7117 LAGRDLLERLATAAGARAGMLQEVVRAQVSVLRLSEGLDVLDAPLTSLGNDLSMLGLE 7176
Qy 144-----143
Db 7177 LRNRIEAVLGITM 7189

RESULT 13
US-08-460-751-2
; Sequence 2, Application US/08460751
; Patent No. 5891628
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen
; APPLICANT: Schneider, Michael
; APPLICANT: Gluckemann, Sandra
; TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
; TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,751
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/413,580
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7638-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4303 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-460-751-2

Query Match 26.9%; Score 208; DB 2; Length 4303;
Best Local Similarity 2.2%; Pred. No. 0.89;
Matches 87; Conservative 19; Mismatches 41; Indels 3862; Gaps 27;

Qy 2 AGG-----1LH-----7
Db 23 AGGPRGCGPCBPCLCGPAPCAACRVNCSGRGLRTLGLPALRIPADATELDVSHNLLRAL 82
Qy 8-----7
Db 83 DVGLLANLSALAEIDISNNKISTLEEGIFANLNLSEINLSGNPFCDGLAWLPQWAE 142
Qy 8-----7
Db 143 QQVVVQPEAATCAGPSLAGOPLLGIPLLDSGCCGEYVACLPDNMSGTVAVVSAAHE 202
Qy 8-----7
Db 203 GLLQPEACSAFCSTGGGLAALSEQWGLCGAAGPSSAPFACLSLCSGPPAPPAPTGRCP 262
Qy 8-----7
Db 263 TLLQHVFPASPGATLVGPHGLASQALAAFHIAAPLPTDTRWDFGSGAEYDAAGPAAS 322
Qy 8-----11
Db 323 HRYVLPGRYHVTAVLALGAGSALLGTDVQVEAPAALELVCPSSVQSDLSLSTONRG 382
Qy 12-----11
Db 383 SGLEAAYSIVALGEBEPARAVHPLCPSPDTEIFPGNGHCYRLVVEKAWLQAOEQCOAWAGA 442
Qy 12-----16
Db 443 ALAMVDSAPVQRFVLSRVTRSLDVMWIGFSTVQGVVEGPAQGEAFSLSCQNWLPGEHP 502
Qy 17-----16
Db 503 ATAEHCVRGLGTGTCNCTDLCSPHYSYCELOPGGPVQDAENLLVCAPSGDLQGLPTPLAQ 562
Qy 17 -DVFOAHQEDTERYV-----LTNLSNIGAEILRDP-----44
Db 563 QDGLSAPHEPVEVMVFPGLRLSREAFLLTAFFGTQELRRPAQLRLOVYVRLSTACTPENG 622
Qy 45-----44
Db 623 SEPESRSPDNRTOLAPACMPGGRWCMPGANICLPDASCHPOACANGCTSGPLPGAYAL 682
Qy 45-----44

Db 683 WREFLFSVPAGPPAQSVTLHGQDVLMLPGDLVGLQHDAGPGALLHCSAPGHPGRAPY 742
QY 45 ----- 44
Db 743 LSNASSMLPHLPAQLEGTMGCPACALRLLAQREQTLVLLGLRPNPGLRLPCRYEVAEV 802
QY 45 ----- 44
Db 803 GNGVSRHNLSCSFDVVSPVAGLRVYPAPRDGRLVYPTNGSALVLQVDSGANATATARP 862
QY 45 --SLGAQR-----VHLVKMV----- 58
Db 863 GGSLSARFENVCALVATVPACPWETNDTLFSVALPMLSGEHHVDDVVVENSASRAWL 922
QY 59 ----- 58
Db 923 SLRVTAEEPICGLRATPSPPEARVLQGLVRYSPVVEAGSDMVFRWTINDKQSLTFQNVVF 982
QY 59 ----- 58
Db 983 NVIYOSAAVKFLSLTASNHSNVTVNVTVERMNRMOGLOVSTVPVAVLSPNATLALTAG 1042
QY 59 ----- 58
Db 1043 VLVDSAVEAFLWTFGDGEQALHQFPYPINESFPVDPDPSVAQVLVEHNVHTHYAAPGEVL 1102
QY 59 ----- 58
Db 1103 LTVLASAFENLTQQVPVSVRASLPSVAVGVSGLVAGRPVTFYPHPLPSGGVLYTWD 1162
QY 59 ----- 58
Db 1163 FGDSPVLTSQPAANHVTYASRGTYHVRLEVNNTVSGAAQADVRVFEELRGLSDMSLA 1222
QY 59 ----- 64
Db 1223 VEQGAPVVVAAVQTGDNITWTGDMGDGTVLSGPEATVEHVYLRQNCVTTVGAGSPAGH 1282
QY 65 -----GA 66
Db 1283 LARSLHLVFLVLELVRPEAACIPTOPDARLTAYVTGNPAHYLFDWTFDGSNTTVRC 1342
QY 67 PNITANLTSS-----LLSVC----- 81
Db 1343 PTVTHNFRSGTFPLALVLSRVNRHAHYFTSICVEPEVGNVTLOPERQFVLGDEAWLVA 1402
QY 82 ----- 81
Db 1403 CAMPPPPYRYTWDGTEBAAPTRARGPEVTFIYRDPGSYLVTVTASNISANDSALVEV 1462
QY 82 ----- 81
Db 1463 QSPVLVTSIKVNGSLGLELQQPYLPSAVGRGPASVYLDLGDGWLGEPEVTHAYNSTGD 1522
QY 82 -----SQT----- 87
Db 1523 FTVRVAGMNEVRSEAWLNVTVKRRVRLVNAASRTVPPLNGSVSFTSLEAGSDVRYSW 1582
QY 88 ----- 87
Db 1583 VLCDRCTPIPGGPTISYTFRSVGTENIIVTAENEVGAQDSIFVYVLQIEGLQVVGGR 1642
QY 88 ----- 87
Db 1643 YFPTNHTVQLQAVRDGTNVSVSWTAWDRGPALAGSGKGFSLTVLEAGTYHVQLRATNM 1702
QY 88 ----- 87
Db 1703 LGSAMADCTMDVFEVPVGLMWAAASPNAAVNTSVTLAEAGSGGVYVYTWLSLEGLSMET 1762
QY 88 ----- 87

Db 1763 SEPTTHSFPTPGLHLVTMTAGNPLGNSANATVEVDVQVPVSGLSIRASEPGGSFVAAGSS 1822
QY 88 ----- 87
Db 1823 VFWGQLATGTVNSWCWAVPGSSKRGPHVTMFPFDAGTFSIRLNASNAVSVSATYNLT 1882
QY 88 -----NPE----- 90
Db 1883 AEEPVLGLVLWASSKVAPGOLVHFQIILAAAGSAVTFRLOVGGANPEVLPQPRFSHSPR 1942
QY 91 ----- 90
Db 1943 VGDHVSVRGKHVHWAQAQVRIVVLEAVSGLQVPNCCEPGIATGTERNFARVORGSRV 2002
QY 91 ----- 90
Db 2003 AYAWYFSLQKVQGDLSVLISGRDVTYTPVAAAGLLEIQVRAFNALGSENRTLVLVEQDAVQ 2062
QY 91 -----DDTD-----PG----- 96
Db 2063 YVALQSGPCFTNRSQAQFEAATSPSPRRVAYHWDGDSGGQDTPDEPRAEHSYLRPGDYRV 2122
QY 97 -----HADL-----VLYI 104
Db 2123 QVNASNLVSFFVAQATVTVQVLACREPEVDVVLPLOVLMRRSQRNYLEAHVDLRDCVTYQ 2182
QY 105 TR----- 106
Db 2183 TEYRWEVYRTASCORPGPARVALPGVDVSRPRLVLPRLALPVGHYCFVVFVSGDTPLT 2242
QY 107 -----PDELPDG----- 114
Db 2243 QSIQANVTVAERLVPPIEGGSYRVMSDTRDLVLDGSESYDNLEDGQOTPLSFHWACVA 2302
QY 115 -----NRQV----- 118
Db 2303 STOREAGCALNFGPRGSSTVTIPRERLAAGVEYTFSLVTWKAGRKEATQTVLIRSGR 2362
QY 119 ----- 118
Db 2363 VPIVSECVCKAQAIVEVRSYVYLEGRCLNCSGSKRGKRWAAFTFSNKTVLDETTT 2422
QY 119 -----RGVTO-----LGAC----- 128
Db 2423 STGSAGMRLVLRGVLRDGEGYTFTLTVLGRSGEBECCASIRLSPNRPLGSCRPLPG 2482
QY 129 ----- 128
Db 2483 AVHALTTKVHPECTGWHDAEDAGAPLVYALLRRRCQGHCEFCVYKGLSSSYGAVLPPG 2542
QY 129 ----- 128
Db 2543 FRPHFVGLVAVVQDOLGAADVVALNRSIAITLPEPNGSATGLTVMLHGLTASVLPGLLRQ 2602
QY 129 ----- 128
Db 2603 ADPOHVIEYSLALVTVLNEYERALDVAAABPKHQHRAQIRKNITETVLSLRVHTVDDIQ 2662
QY 129 -----SPT----- 131
Db 2663 QIAAALAQCMGSPRELVCRCCLKQTLHKLEAMMLILOAETTAGTVTPTAIGDSILNITGD 2722
QY 132 ----- 131
Db 2723 LIHLASSDVRAPQPSSELGAESPRMVASQAYNLTSLMRILMRSRVLENEPLTLAGEEIV 2782
QY 132 ----- 131
Db 2783 AOGKRSDRSLCYGGAPGPCCHFSIPEAFSGALANLSDVVQLFLVDSNPPFPFGYISNY 2842
QY 132 ----- 131
Db 2843 TVSTKVSMAFQTQGAQIPIERLASERAITVKVPNNSDWAARGHRSSANSANSVVVQPO 2902

```
QY 132 ----- 131
Db 2903 ASVCVVTLSSNPAAGLHLQNLNYTLDDGHYLSEEPYLAIVLHSEPRNEHNCASRR 2962
QY 132 -----WS----- 133
Db 2963 IRPESLOQADHRPYTFIFSPGSRDPAGSYHLNLSHFRWSALQVSLYTSLCQYFSEED 3022
QY 134 ----- 133
Db 3023 MVRTEGLLPLEETSPRQAVCLTRHLTAFCASLFVPPSHRVFVPEPTADVNVIVMLTCA 3082
QY 134 -CLIT----- 137
Db 3083 VCLVTYVMAAILHLKLDLDASRRAIPFCGQRGRPKYEILVKTGRCGSTTAHVGMIL 3142
QY 138 ----- 137
Db 3143 YGVDSRSGRHLDDGDFRHNLSUIFRIATPHSLGVMKIRVMDNKGSLSPAWFLQHVIV 3202
QY 138 ----- 137
Db 3203 RLOQTARSAFFLVNDWLSVETEANGGLVEKEVLAASDAALLRFRLLVAELQGRFFDKHI 3262
QY 138 ----- 137
Db 3263 WLSINDRPPRSRFRTRIORATCCVLLICLFLGANAVVYGVGDSAYSTGHVSRLSPLSVDT 3322
QY 138 ----- 137
Db 3323 VAVGLVSVVVPVYLAILFLFRMSRSKVAGSPPTPAGQQVLDIDSCLDSSVLDSSFLT 3382
QY 138 ----- 137
Db 3383 FSGLHAEQAFVQMKSDLFLDPSKSLVCPWSPGEGTLWPDLSDFPSIVGSNLRQLARGOA 3442
QY 138 ----- 137
Db 3443 GHGLGPEEDGFSLASPYSPAKSFSASDEDLIOQVLAEGVSPAPTQDTHMETDLSLSS 3502
QY 138 ----- 137
Db 3503 TPCEKTETLQRLGELGPPSPCLNWEQQAARLSTGLVEGLRKLPAWCASLAHGLS 3562
QY 138 ----- 137
Db 3563 LLLVAVAVSGWVGASFPFGVSVAMLLSSSASFLASFLGWEPKLVLEALYFSLVAKRL 3622
QY 138 -EO----- 139
Db 3623 HPDEDDTLVESPAVTPVSARVPRVRPPHGFALFLAKEARKVKRLHGMRLSLLVYMLFLL 3682
QY 140 ----- 139
Db 3683 VTLASYGDSCHGHAYRLQSAIKQELHSRAFLAITRSEELWPMMAHVLLPYVHGNQSSP 3742
QY 140 ----- 139
Db 3743 ELGPPRLROVRLQEALYDPDPPGRVHTCSAAGGFSTSDYDVGWESPHNGSGTWAYSAPDL 3802
QY 140 ----- 139
Db 3803 LGANWGCAYVDSGGYVQELGSLSESRDLRFLQLHNWLDNRSRAVFELETRYSPAVG 3862
QY 140 ----- 139
Db 3863 LHAAVTLRLEFPAGRALAALSVPFALRRLSAGLSPLLTSTVCLLLFVAVHFAVAEARTW 3922
QY 140 -----TGFD----- 143
Db 3923 HREGRWRLRGAMARWLLVALTAATALVRLAQLGAADROWTRFVRGRPRRFTSPDQVAH 3982
```

```
QY 144 -----LGVTI 148
Db 3983 VSSAARGLAASLLFLLLVKAAQHVRPVRQMSVFGKTLCRALPELLGVTL 4031
RESULT 14
US-09-320-878-1
; Sequence 1, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE OF INVENTION: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320.878A
; EARLIER FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 1
; LENGTH: 4551
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-1
Query Match 26.8%; Score 207; DB 3; Length 4551;
Best Local Similarity 1.9%; Pred. No. 1.2;
Matches 84; Conservative 19; Mismatches 42; Indels 4181; Gaps 23;
QY 1 AAGG----- 4
Db 52 AAGGQAVTDVPADRWNAGDFYDPRDSAPGRSNRWGGFIEDVDRFDAAFGISPREAEM 111
QY 5 ----- 4
Db 112 DPOORLALGWEALERAGIDPSSLTGTRTGVFAGAIWDDYATLKHROGGAITPHTVTG 171
QY 5 ----- 4
Db 172 LHRGIIANRLSYTLGLRPSMWVDSSGSSSLVAHVLACESLRRGESELALAGVSLNLP 231
QY 5 ----- 4
Db 232 DSIIGASKFGLSPDGRAYTFDARANGYVRGEGGFVVKRLSRVADGDPVLAVIRGSA 291
QY 5 ----- 4
Db 292 VNNGAAQGWTTTDAQOAEAVLREAHHERAGTAPADVRYVELHGTGTVPDPIEAALGAA 351
QY 5 -----ILHLE----- 9
Db 352 LGTRPAGQPLLGVSVKTNIGHLEGAAGIAGLIKAVLAVRGRALPASLNYETPNPAIPFE 411
QY 10 ----- 9
Db 412 ELNLRVNTYLPWEPEHGDGMVMVGVSSFGMGGTNAHVLEAPGVVVGASVVESTVGGS 471
QY 10 ----- 9
```

```
Db 472 AVGGVVPVWVSAKSAALDAQIERLAAFAASDRDTDGDAGAVDAGAVARVLACG 531
Qy 10 -----LLVAVGPD----- 17
Db 532 RAQFEHRAVVGSGPDDLAAALAAPEGLVRGVASGVRVAFVFPQGCTQWAGGAELDS 591
Qy 18 --VFOA----- 21
Db 592 SAVFAAAMAECEAALSPLYVDWSLEAVVRQAPGAPTLERVDVQVPTFAVMVSLARVQHH 651
Qy 22 ----- 21
Db 652 GVTPOAVVGHSGEIAAAVAGALSDDAARVVTLRKSIAAHLACKGKMLSALSSEDV 711
Qy 22 ----- 21
Db 712 LERLAGFDGLSVAANGPTATVVSGDPVQIEELARACEADGVRARVIPDYASHSRQVEI 771
Qy 22 ----- 21
Db 772 IESELAEVLAGLSPOAPRVPPFFSTLEGAWITEPVLDDGGYWRNLRHVRGFAPAVETLATD 831
Qy 22 ----- 21
Db 832 EGFTHFEVSAHPVLTMLPGTVTGLATLRDNGGODRLVASLAEWANGLAVDWSPLP 891
Qy 22 ---HOED-----TER----- 28
Db 892 SATGHSDLPTYAFOTERHWLGEIEALAPAGEPAVQPAVLRTAAAPAELODEQLRVIL 951
Qy 29 ----- 28
Db 952 DKVRAQTAQVLGYATGGQIEVDTRPREAGCTSLTGVDLRNRINAAFGVRMAPSMIFDPPT 1011
Qy 29 ----- 28
Db 1012 PEALAEQLLVVHGEAANPAGAPAPVAAAGAVDEPVAIVGMACRLPGVASPEDLWRL 1071
Qy 29 ----- 28
Db 1072 VAGGDAISEFPQDRGWDEGLYHPDPEHPGTSYVRQGGFIENVAGFDAFFGISPREAL 1131
Qy 29 -----VULT 32
Db 1132 AMDPOORLLLETWEAVEADAGIDPTSLRGRQVGVFTGANTHEYGPSLRDGGEGDGLYLLT 1191
Qy 33 N-----LNIGAEELR----- 42
Db 1192 GNTASVMSGRVSYTLGLEGPALTVDTACSSSLVALHLAVQALRKEVDMALGAVVAMPT 1251
Qy 43 ----- 42
Db 1252 PGMFEFSRQRLAGDGRSKAFASADGTSWSEGVVLLVERLSDARRNGHQVLAVRGS 1311
Qy 43 ----- 42
Db 1312 AVNQDGASNGLTAPNGPSQQRVIRRALADARLTTSDDVDVVEAHGTGRLGDPIDAEQAALIA 1371
Qy 43 ----- 42
Db 1372 TYGQRDDEQPLRLGSLKSNIGHTQAAAAGVSGVIKWQAMRHGLLPKTLHVDEPSDQIDW 1431
Qy 43 -----DPSLG- 47
Db 1432 SAGAVELLTEAVDWPEKODGRLRAAVSSFGISGTAHVVLEAPVWVEGASVVPVPSVG 1491
Qy 48 -----AQF 50
Db 1492 SAVGGVTPVWVSAKSAALDAQIERLAAFAASDRDTDDADAGAVDAGAVAHVLADGRAQF 1551
Qy 51 ----- 50
Db 1552 EHRAVALGAGADDLVQALADPDGLIRGTASGVGRVAFVFPQGCTQWAGGAELDSSAVF 1611
```

```
Qy 51 -----RVHLV----- 55
Db 1612 AAAMAECEAALSPLYVDWSLEAVVRQAPGAPTLERVDVQVPTFAVMVSLARVQHHGVT 1671
Qy 56 ----- 55
Db 1672 QAVVGHSGEIAAAVAGALPLDDAARVVTLRKSIAAHLACKGKMLSALNEDAVLERL 1731
Qy 56 ----- 55
Db 1732 SDFDGLSVAANGPTATVVSGDPVQIEELARACKADGFRARIIPVDYASHSRQVEIESE 1791
Qy 56 ----- 55
Db 1792 LAOVLAGLSPOAPRVPPFFSTLEGWITEPVLDDGTWYRNLRHVRGFAPAIETLAVDEGFT 1851
Qy 56 ----- 55
Db 1852 HFVEVSAHPVLTMTLPETVTGLTLRREOGGOERLVTSLAEAWNGVLPVAMTSLPATAS 1911
Qy 56 ----- 55
Db 1912 RPGLPTYAFOERYWLENTPAALATGDDWRYRIDWKELPAEAGSERTGLSGRWLAVTPE 1971
Qy 56 ----- 55
Db 1972 HSAQAAVLTALVDAGAKVEVLTAGADDDREALAARLTALTGTGFTGVVSLDGLVPQV 2031
Qy 56 -----KMWILTEPE----- 64
Db 2032 AMVOALGDAGIKAPLMSVTOGAVSVGRGLDTPADPDRAMLGLGRVVALEHPERWAGLVLD 2091
Qy 65 ----- 64
Db 2092 PAQPDAAALAHVLTALSGATGEDQIAIRTTGLHARRLARAPLHGRRTDRMQPHGTVLIT 2151
Qy 65 -----GAPNITANLTSS----- 76
Db 2152 GGTGALGSHAARMAWHGAHEHLVLSRSGEOAPGATQLTAEALTASGARVITAAACDVADPH 2211
Qy 77 ----- 76
Db 2212 AMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTASOVRRRAHRAKAVGASVLDDELTRDLD 2271
Qy 77 ----- 76
Db 2272 LDAFVLFSSVSTLGIPIGOGNYAPHNAYLDALAARRRATCRSAVSANGPWGCGMAAGD 2331
Qy 77 -----LLSVCGWSQTI----- 87
Db 2332 GVAERLRNHGVPGMDPELAALESALGRDETAITVADIDWDRFYLAYSSGRPOPLVEEL 2391
Qy 88 ----- 87
Db 2392 PEVRIIDARDSATSCGGSSAQGANPLAERLAAAAAPGERTELLGLVRAQAAVLRMRS 2451
Qy 88 ----- 87
Db 2452 PEDVAADRAFKDIGFDSLAVAGVLRNRLTRATGLQLPATLVFDHPTPLALVSLRSEFLGD 2511
Qy 88 -----NPED----- 91
Db 2512 EETADARRSAALPATVAGAGAGAGTDDADDPIAIVAMSCRYPGDIRSPEDLWMLSEGG 2571
Qy 92 -----DTDP----- 95
Db 2572 EGITPPPTDRGWLDGLYDADPDALGRAYVREGGFLHDAAEFDAEPFGVSPREALMDPO 2631
Qy 96 ----- 95
Db 2632 QRMLLTTSWEAFERAGIEPASLRSGSSTGVFGLSYDYAARVNPAPRAGVEGYLLTGSTPS 2691
```

```
QY 96 ----- 95
Db 2692 VASGRIAYTFLEGPATTVDTCSSILTAHLAVRALRSGECTMALAGGVAMMATPHFV 2751
QY 96 ----- 95
Db 2752 EFSQRALAPGRSKAFSADADGFAAEGVGLLVERLSDARRNGHPVLAVVGTAVNQD 2811
QY 96 ----- 95
Db 2812 GASNGLTAPNGPSOORVIRQALADARLAPGDI DAVETHGTGTSIGDPIEAQLOATYGKE 2871
QY 96 ----- 95
Db 2872 RPAERPLAIGSVKSNIGHTOAAAGAAGIIMVLAMRHGTLPKTLHADEPSPHVDWANSGL 2931
QY 96 ----- 95
Db 2932 ALVTEPIDWAGTGPRAAVSSFGISGTNAHVLEQAPDAAGEVLGADEVSEVSETVAMA 2991
QY 96 ----- 95
Db 2992 GTACTSEAESEASEAPAPGSPREASLPGLPHVLSAKDEQSLRGQAAALHAWLSEPA 3051
QY 96 ----- 95
Db 3052 DLSADGPARLDVGYTLATSRFAFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTA 3111
QY 96 ----- 102
Db 3112 RDGTAFLEFTGQSGRPGAGRELYDRHPVFAALDEICAHLDGHLLEPLLDVNFMAEGSA 3171
QY 103 ----- 102
Db 3172 EAALLDETRYTOCALFALEVALFRLVESWGRPAALLGHSVGETAAAHVAGVFSLADAAR 3231
QY 103 ----- 108
Db 3232 LVAARGRLMOELPAGGAMLAVOAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSGDAD 3291
QY 109 ----- 108
Db 3292 AAREAEAYWSGLRRTRALRVSHAFHSAHMDGMLDGPRAVLETFEPRRSLTVVSNVTGL 3351
QY 109 ----- 108
Db 3352 AAGPDDLCDEPYWVRHVGTVRFLDGYRVLRLGVRTCLLELPGDGVLTANAADGLADTPA 3411
QY 109 ----- 108
Db 3412 DSAAGSPVGPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWH 3471
QY 109 ----- 113
Db 3472 ANFAGSAHRVDLPTYSFRDRYWLDPADAADTAVDTAGLGGTADHPFLGAVVSLPDRDG 3531
QY 114 ----- 124
Db 3532 LLLTGRLSLRTHPLADHAVLGSVLLPGAAVVELAAHAAESAGLRDVRLETLLEPLVLP 3591
QY 125 -GGA- -CSPWSC- -134
Db 3592 HGGVELRVTVGAPAGEPGESAGDGPVSLHSRLADAPAGTAWSCHATGLLATDRPELP 3651
QY 135 ----- 134
Db 3652 VAPDRAAMWPPQAEVPLDGLYERLDGNGLAFPLFQGLNAVVRVEGEVFADIALPATT 3711
QY 135 ----- 134
Db 3712 NATAPATANGGSAAPYGIHPALLDASLHATAVGLGVDEPELVRVPFWHSQVTVHAAG 3771
QY 135 ----- 134
```

```
Db 3772 AAAARVRLASAGTDAVSLSLTGDGRPLVSVRLTLRPTVTDQAAASRVGGLMHRVAMRP 3831
QY 135 ----- 134
Db 3832 YALASSEODPHATSYGPTAVLGKDELKVAALASAGVEGLYPLDLAALSQDVAAGAPAP 3891
QY 135 ----- 134
Db 3892 RTVLAPLPAGPADGGAEGVGTAVRTLELLOAMLADEHLAGTRLLLVTRCAVRDPESGA 3951
QY 135 ----- 134
Db 3952 DGGEDLSHAAAMGLVRTAQTENPGRFGLDLADASSVRYTLPSVLSDAGLRDEPOLALH 4011
QY 135 ----- 141
Db 4012 DGTIRLARLASVRPETGTAAPALAPEGTVLLTGGTGGGLGVARHVGVGVRRLLLVSR 4071
QY 142 ----- 141
Db 4072 RCTDAPGADELVHELEALGADVSAACDVADREALTAVLDAIPAHEHPLTAVVHTAGVLSD 4131
QY 142 ----- 141
Db 4132 GTLPSTMTEDVEHVRPKVDAAFLDLDELSTPAYDLAAFMFSSAAAAMVFGAGOGAYAAA 4191
QY 142 ----- 141
Db 4192 NATLDALAWRRRAAGLPALSLGWLWAE'TSGMTGELGOADLRMRSRAGIGGISAEGIAL 4251
QY 142 ----- 141
Db 4252 LDAAALRDDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRRAPSAASATTA 4311
QY 142 ----- 141
Db 4312 GTAGTGTADGAAETAATAVTLADRAATVDGPARQRLLEFVVGVEAEVLGHARGHIDAER 4371
QY 142 --FDLG 145
Db 4372 GFLLDLG 4377
```

RESULT 15

```
US-09-105-537-31
; Sequence 31, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105.537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 4613
; TYPE: PRN
; ORGANISM: Streptomyces venezuelae
US-09-105-537-31
```

Query Match 26.8%; Score 207; DB 4; Length 4613;
Best Local Similarity 1.9%; Pred. NO. 1.2;

Matches 84; Conservative 19; Mismatches 42; Indels 4181; Gaps 23;

QY 1 AAGG----- 4

Db 114 AAGGQAVTDPADRNAGDFYDPRAPSGRNSRMGGFIEDVDRFDAAFFGISPREAAEM 173

```
QY 5 ----- 4
Db 174 DPQRLALELGMERALERAGIDPSSLTGRTRGVFAGAIWDDYATLKHROGAAITPHTVTG 233
QY 5 ----- 4
Db 234 LHRGITANRLSYTLGURGSMVVDSCQSSSLVAVHLACESLRGESELALAGVSNLVP 293
QY 5 ----- 4
Db 294 DSIIGASKFGLSPDCRAVTFDARANGYVRGEGGFWLKRLSRAVDGDPVLAVIRGSA 353
QY 5 ----- 4
Db 354 VNNGGAAQGWTFDAAQAEVILBEAHERAGTAPADVRYVELHGTGTPVGDPTEAAALGAA 413
QY 5 ----- 9
Db 414 LGTRPAGQPLLVGSVKTNIGHLEGAAGIAGLIKAVLAVGRALPASLNYETPNPAIPFE 473
QY 10 ----- 9
Db 474 ELNLRVNTYELPWEPEHDCQRMVGVSSFGMGGTNAHVLEEAPOGRCRGASVVESTVGS 533
QY 10 ----- 9
Db 534 AVGGGVVPMVWSAKSAAALDAQIERLAFAFASDRDRTGDVADAGAVDAGAVARVLGG 593
QY 10 ----- 17
Db 594 RAQFEHRAVVVSGPDDLAAALAAPEGLVRGVASGVRVAFVFPQGTQWAGMGAELDS 653
QY 18 --VFOA----- 21
Db 654 SAVFAAANAACEAALSPYVDLSLEAVVROAPGAPTLERVDVVQVPTFAMVSLARVMOQH 713
QY 22 ----- 21
Db 714 GVTPOAVVCHSOGETAAYVAGALSDDMARVVTLRSKSIAAHLAKGKGMLSLALSEDAV 773
QY 22 ----- 21
Db 774 LERLAGFDGLSVAANGPTATVVSGDPVQIEELARACEADGVRARVIPVDYASHSRQVEI 833
QY 22 ----- 21
Db 834 IESELAEVLAGLSPOAPRVPPFFSTLEGAWITEPVLDDGYWYRNLRRHRYGFAPAVETLATD 893
QY 22 ----- 21
Db 894 EGFTHFVEVSAHPVLTMLPGTVTGLATLRRDNGGQDRVLVASLAEAWANGLAVDWSPLLP 953
QY 22 --HOED-----TER----- 28
Db 954 SATGHSDLPYAFQTERHWLGEIEALAPAGEPAVQPAVLRTAAEPAELDRDEQLRVIL 1013
QY 29 ----- 28
Db 1014 DKVRAQTAQVLGYATGGQIEVDRTFREAGCTSLTGVDLRNRINAAFGVRMAPSMIFDPPT 1073
QY 29 ----- 28
Db 1074 PEALAEQLLVVHGEAANPAGAEPAVAAAGAVDEPAIVCMACRLPGGVASPEDLWRL 1133
QY 29 ----- 28
Db 1134 VAGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYVROGGFIENVAGFDAFFGISPREAL 1193
QY 29 -----VYLT 32
Db 1194 AMDPOORLLLETSWAEVADAGIDPTSLRGRQVGVGTGANTHEYGPSLRDGGEGDGLYLLT 1253
QY 33 N-----LNIGAELLR----- 42
```

```
Db 1254 CNTASVMSGRVSYTLGLEPALTVDTACSSSLVALHVAQALRRKGEVDMALAGGVAVMPT 1313
QY 43 ----- 42
Db 1314 PGMFVFSRQRLAGDGRSKAFAASADGTSWSEGVGVLLVERLSDARRNGHOVLAVRGS 1373
QY 43 ----- 42
Db 1374 ALNQDGASNGLTAPNGPSQORVIRRALADARLTSTDVDVVEAHGTGRLGDPTEAQAALIA 1433
QY 43 ----- 42
Db 1434 TYGGRDDEOPLRLGSLKSNIGHTQAAAGVSGVJMWQAMRHGLLTKTLHVDEPSDDIDW 1493
QY 43 -----DPSLG- 47
Db 1494 SAGAVELLTEAVDWPEKQDGLRRAAVSSFGISCTNAHVLEEAAPVVVEGASVVEPSVGG 1553
QY 48 -----AOF 50
Db 1554 SAVGGVTPMWVWSAKSAAALDAQIERLAFAFASDRDRTDDADAGAVDAGAVHVLADGRAOF 1613
QY 51 ----- 50
Db 1614 EHRAVALGAGADLLVQALADPDGLIRCTASGVRVAFVFPQGTQWAGMGAELLDSSAVF 1673
QY 51 -----RVHLV----- 55
Db 1674 AAAMEACEAALSPPVDWSLEAVVRQAPGAPTLERVDVVQVPTFAMVSLARVMOHGVTP 1733
QY 56 ----- 55
Db 1734 QAVVHSGEIAAAYVAGALPLDDAARVVTLRSKSIAAHLAKGKGMLSLALNEDAVLERL 1793
QY 56 ----- 55
Db 1794 SDFDGLSVAANGPTATVVSGDPVQIEELAQACKADGFRARIIPVDYASHSRQVEIIESE 1853
QY 56 ----- 55
Db 1854 LAQVLAGLSPOAPRVPPFFSTLEGTWITEPVLDDGYWYRNLRRHRYGFAPAIETLAVDEGPT 1913
QY 56 ----- 55
Db 1914 HFVEVSAHPVLTWLTPEVTGLTGLRREGQGERLVTSLABAWWNGLPVAVNTSLLPATAS 1973
QY 56 ----- 55
Db 1974 RPLGPTYAFOERYWLENTPAALATGDDWRYRIDWKRLPAEGSERTGLSERWLAVTPEP 2033
QY 56 ----- 55
Db 2034 HSAQAAAVLTALVDAGAKVEVLTAGADDDREALAARLTALTGGDGTGVVSLDGLVPQV 2093
QY 56 -----KMVLTEPE----- 64
Db 2094 ANVQALGDAGIKAPLWSVTOGAVSVGRDLTPADPDRAMLWGLGRVVALEHPERWAGLVDL 2153
QY 65 ----- 64
Db 2154 PAQPDAAALAHVLTALSGATGEDQIAIRTTGLHARRLARAPLHGRPRTRDWOPHGTVLIT 2213
QY 65 -----GAPNITANLTSS----- 76
Db 2214 GGTGALGSHAARMAWHGAHGLLLVSRSGEQAPGATQLTAELTASGARVTIAACDVADPH 2273
QY 77 ----- 76
Db 2274 AMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRRAHRAKAVGASVLDELTRDLD 2333
QY 77 ----- 76
```

```
Db 2334 LDAFVLSSVSTLIGPGQNYAPHNAYLDALARRRATGRSAVSAMGPMWGGMAAGD 2393
Qy 77 -----LLSVCGWSQTI----- 87
Db 2394 GVAERLRNHGVGMPELALAESALGRDETAITVADIDWDRFYLAYSSGRPOPLVEEL 2453
Qy 88 ----- 87
Db 2454 PEVRIIDARDSATSGOGSSAQOANPLAERLAAAPGERTEILLGLVRAQAAVLRMS 2513
Qy 88 ----- 87
Db 2514 PEDVAADRAFKDIGFDSLACVLRNLTRATGLQLPATLVDFHPTPLALVLSLRSEFLGD 2573
Qy 88 -----NPED----- 91
Db 2574 EETADARRSAALPATVAGAGAGAGTADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGG 2633
Qy 92 -----DTDP----- 95
Db 2634 EGITPFTDRGWLGLYDADPDALGRAYVREGGFLHDAAEFDABEFGVSPREALAMDQ 2693
Qy 96 ----- 95
Db 2694 QRMLLTTSWEAFERAGIEPASLRGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPS 2753
Qy 96 ----- 95
Db 2754 VASGRIAYTFGLEGPATTVDTCSSSLTALHLAVRALRSGECTMALAGGVAMTAPHMFV 2813
Qy 96 ----- 95
Db 2814 EFSQRALAPDGRSKAFSADADFGAAEGVGLLVERLSDARRNGHPVLAVVRGTAVNQD 2873
Qy 96 ----- 95
Db 2874 GASNGLTAPNGSQORVIRQALADARLAPGDIIDAVETHGTGTSGLDPIEAQGLQATYKGE 2933
Qy 96 ----- 95
Db 2934 RPAERPLAIGSVKSNIGHTQAAAGAAIIMVLMRHGTLPKTLHADESPHVDWANSGL 2993
Qy 96 ----- 95
Db 2994 ALVTEPIDWPGTGPRAAVSFGISGTNAHVLEQAPDAAGEVLGADEVPEVSETVAMA 3053
Qy 96 ----- 95
Db 3054 GTACTSEVAGSEASEAPAPGCSREASLPCHLPWVLSAKDEQSLRGQAAALHAWLSEPA 3113
Qy 96 ----- 95
Db 3114 DLSADGPARLRDVGTYLTATSRFAHRAAVTAADRDGDLGLATLAQGGTSAHVHLOTA 3173
Qy 96 -----GHADIVL----- 102
Db 3174 RDGTTAFUFTGGSORPGAGRELYDRHPVFARALDEICAHLDGHLELPLLDVMAFAGESA 3233
Qy 103 ----- 102
Db 3234 EAALLDETRYTCALFALEVALFRLVESWGMRPALLGHSVGETIAAAHVAGVFSLADAAR 3293
Qy 103 -----YITRFD----- 108
Db 3294 LVAARGRLMQELPAGGAMLVAQAADEIRVMLETEERYAGRLDVAAVNGPEAAVLSDGAD 3353
Qy 109 ----- 108
Db 3354 AAREAEYWSGLRTRRALRVSHAFSAHMDGMLDGFRAVLETFEVRFRPSLTIVVSNVTGL 3413
Qy 109 ----- 108
Db 3414 AAGPDDLCDPEYVWRHVRGTVRFILDGVRVLRDLGVRTCLGLPGDGLVTAMAADGLADTPA 3473
```

Search completed: March 20, 2003, 12:41:25
Job time : 149 secs

```
Qy 109 ----- 108
Db 3474 DSAAGSPVCSPAGSPADSNAGALRPRPLLALLRRRSETETVADALGRAHANGTGPDPWH 3533
Qy 109 -----LELPD----- 113
Db 3534 ANFAGSAGHRVDLPTYSFRRDRYWLDAPAADTAVDTAGLGLCTADHPLLGAVSVLPDRDG 3593
Qy 114 -----GNQVRGVGTOL----- 124
Db 3594 LLLTGLRLSLRTHPWADHAVGLSVLLPGAAMVELAAHAAESAGLRDRVRELTLLLEPLVLP 3653
Qy 125 -CGA-----CSPTWSC----- 134
Db 3654 HGVVELRVTVGAPAGEGEGSAGDGPVSLHSRLADAPAGTAMWSCHATGLLATDRPELP 3713
Qy 135 ----- 134
Db 3714 VAPDRAAMWPQGAEEVPLDGLYERLDGNGLAFLGFLFQGLNVMRYEVEGFADIALPATT 3773
Qy 135 ----- 134
Db 3774 NATAPATANGGSSAAAAPYGIHPALLDASLHAIAGVGLVDBPELVRVPFHMVGTVHAAG 3833
Qy 135 ----- 134
Db 3834 AAAARVRLASAGTDAVLSLTDGGRPLVSVVERLTLPVTADQAAASRVGGLMHRVAVRP 3893
Qy 135 ----- 134
Db 3894 YALASSGEQDPHATSYGPTAVLGKDELKVAALLESAGVEVGLYPDLAALSODVAAGAPAP 3953
Qy 135 ----- 134
Db 3954 RTVLAPLPAGPADGGAEGVGTAVRTLELQAWLADEHLACTRLLLVTRGAVRDPGSGA 4013
Qy 135 ----- 134
Db 4014 DGGEDLSHAAAWGLVRTAQTENPGRFGLLDLADDASSYRTLPSVSLDAGLRDEPQLALH 4073
Qy 135 -----LITEDTG----- 141
Db 4074 DGTIRLARLASVRPETGTAAPALAPETGVLLTGTGTGGLVLRHVGVGVRRLLLVSR 4133
Qy 142 ----- 141
Db 4134 RGTDPAGADELVHELEALGADSVVAACDVADREALTAVLDAI PAEHPLTAVVHTAGVLS 4193
Qy 142 ----- 141
Db 4194 GTLPSMTTEDVEHVLRPKVDAAFLLDLTSTPAYDLAAFMVFSSAAAVFGAGOGAYAAA 4253
Qy 142 ----- 141
Db 4254 NATLDALAMRRRAAGLPALSLGMLWAETSGMTGELQOALRRMSRAGIGGISDAEGIAL 4313
Qy 142 ----- 141
Db 4314 LDAALRRDRHPVLLPLRLDAAGLRDAAGNDPAGI PALFRDVGARTVVRARPSAASATTA 4373
Qy 142 ----- 141
Db 4374 GTAGTCTAGAAETAATVTLADRAATVDGPARQORLLLEFFVGVGVAEVLGHARGHIDAER 4433
Qy 142 --FDLG 145
Db 4434 GFLDLG 4439
```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:37:58 ; Search time 18 Seconds
(without alignments)
439.578 Million cell updates/sec

Title: SEQID_1_15FUSED

Perfect score: 773

Sequence: 1 AAGGILHLELLVAVGPDVFQ.....SPTWSCLITGTGDLGVIT 148

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.0

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 221153

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA.*
- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/FCI_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/FCIUS_PUBCOMB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	773	100.0	148	US-09-833-328-4	Sequence 4, Appli
2	773	100.0	150	US-09-833-328-6	Sequence 6, Appli
3	773	100.0	242	US-09-978-979-1	Sequence 1, Appli
4	773	100.0	1416	US-09-836-712-2	Sequence 2, Appli
5	770	99.6	203	US-09-781-0808-2	Sequence 2, Appli
6	770	99.6	1120	US-09-781-0808-11	Sequence 11, Appli
7	716	92.6	136	US-09-833-328-15	Sequence 15, Appli
8	702	90.8	133	US-09-833-328-2	Sequence 2, Appli
9	272	35.2	26926	US-09-759-508B-2	Sequence 7, Appli
10	237	30.7	997	US-09-918-171A-7	Sequence 7, Appli
11	235	30.4	11877	US-09-860-846-5	Sequence 6, Appli
12	235	30.4	11877	US-09-861-289-6	Sequence 6, Appli
13	231	29.9	1690	US-09-788-043C-5	Sequence 5, Appli
14	229	29.6	2150	US-09-321-987B-2	Sequence 2, Appli
15	229	29.6	2165	US-09-800-729-155	Sequence 155, App
16	228	29.5	7968	US-10-077-130-5	Sequence 5, Appli
17	222	28.7	730	US-10-163-547-16	Sequence 16, Appli
18	220	28.5	6304	US-10-147-026-16	Sequence 16, Appli
19	215	27.8	4349	US-10-160-758-15	Sequence 15, Appli

ALIGNMENTS

RESULT 1

US-09-833-328-4 Application US/09833328

Patent No. US20020136713A1

GENERAL INFORMATION:

APPLICANT: Laemmle, Bernhard

APPLICANT: Schwarz, Hans-Peter

APPLICANT: Schefflinger, Friedrich

APPLICANT: Antoine, Gerhard

APPLICANT: Kerschbaumer, Rndolf

APPLICANT: Tagliavacca, Luigina

APPLICANT: Zimmermann, Klaus

APPLICANT: Furlan, Miha

APPLICANT: Turecek, Peter

APPLICANT: Gerritsen, Helena E.

TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protea

FILE REFERENCE: 247.00CIP

CURRENT APPLICATION NUMBER: US/09/833,328

PRIOR FILING DATE: 2001-04-12

PRIOR APPLICATION NUMBER: 09/721,254

PRIOR FILING DATE: 2000-11-22

NUMBER OF SEQ ID NOS: 15

SOFTWARE: PatentIn version 3.1

SEQ ID NO 4

LENGTH: 148

TYPE: PRT

ORGANISM: human

US-09-833-328-4

Query Match 100.0%; Score 773; DB 10; Length 148;

Beat Local Similarity 100.0%; Pred. No. 4.4e-31;

Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAGGILHLELLVAVGPDVFQAHQEDTERVLTNLNIGAEILLRDPISLGAQFRVHLVKMVL 60

Db 1 AAGGILHLELLVAVGPDVFQAHQEDTERVLTNLNIGAEILLRDPISLGAQFRVHLVKMVL 60

Oy 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGCHADLVLYITRFDLELPDGNQVRG 120

Db 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGCHADLVLYITRFDLELPDGNQVRG 120


```
QY 121 VTOLGGACPTWSCLITDGTGDLGVTI 148
DB 121 VTOLGGACPTWSCLITDGTGDLGVTI 148

RESULT 2
US-09-833-328-6
; Sequence 6, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Scheiflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Rüdolf
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Gerritsen, Helena E.
; TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Acti
; TITLE OF INVENTION: Comprising a Polypeptide Chain with the Amino Acid Sequence AAGG
; FILE REFERENCE: 247.00CIP
; CURRENT APPLICATION NUMBER: US/09/833.328
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 150
; TYPE: PRT
; ORGANISM: human
; US-09-833-328-6

Query Match 100.0%; Score 773; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 4.6e-31;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 60
DB 3 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 62

QY 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPDGNRQVRG 120
DB 63 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPDGNRQVRG 122

QY 121 VTOLGGACPTWSCLITDGTGDLGVTI 148
DB 123 VTOLGGACPTWSCLITDGTGDLGVTI 150

RESULT 3
US-09-978-979-1
; Sequence 1, Application US/09978979
; Patent No. US20020151702A1
; GENERAL INFORMATION:
; APPLICANT: Racie, Lisa, A.
; APPLICANT: Twine, Natalie, C.
; APPLICANT: Agostino, Michael, J.
; APPLICANT: Wolfman, Neil
; APPLICANT: Morris, Elisabeth
; TITLE OF INVENTION: Aggreganase Molecules
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Home Products Corporation
; STREET: One Campus Drive
; CITY: Parsippany
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/978.979
FILING DATE: 16-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/60/241.469
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Kapinos, Ellen J.
REGISTRATION NUMBER: 32,345
REFERENCE/DOCKET NUMBER: GI 5435p
TELECOMMUNICATION INFORMATION:
TELEPHONE: (973) 660-5000
TELEFAX: (973) 683-4117
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 242 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-978-979-1

Query Match 100.0%; Score 773; DB 10; Length 242;
Best Local Similarity 100.0%; Pred. No. 1.4e-30;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 60
DB 42 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 101

QY 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPDGNRQVRG 120
DB 102 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPDGNRQVRG 161

QY 121 VTOLGGACPTWSCLITDGTGDLGVTI 148
DB 162 VTOLGGACPTWSCLITDGTGDLGVTI 189

RESULT 4
US-09-836-712-2
; Sequence 2, Application US/09836712
; Patent No. US20010049106A1
; GENERAL INFORMATION:
; APPLICANT: PFIZER INC.
; TITLE OF INVENTION: ADAMTS POLYPEPTIDES, NUCLEIC ACIDS ENCODING THEM, AND
; FILE REFERENCE: PC10851A
; CURRENT APPLICATION NUMBER: US/09/836.712
; CURRENT FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1416
; TYPE: PRT
; ORGANISM: Human
; US-09-836-712-2

Query Match 100.0%; Score 773; DB 10; Length 1416;
Best Local Similarity 100.0%; Pred. No. 8.1e-29;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 60
DB 98 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 157

QY 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPDGNRQVRG 120
```

Db 158 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQVRG 217
QY 121 VTOLGGACSPWTSCLITDGTGFDLGVTI 148
Db 218 VTOLGGACSPWTSCLITDGTGFDLGVTI 245

RESULT 5

US-09-781-080B-2
; Sequence 2, Application US/09781080B
; Patent No. US20020142439A1
; GENERAL INFORMATION:
; APPLICANT: HOLLOWAY, JIM
; APPLICANT: SHEPPARD, PAUL
; APPLICANT: YAMAMOTO, GAYLE
; TITLE OF INVENTION: Anti-Angiogenic Intestinal Peptides,
; TITLE OF INVENTION: zdn125
; FILE REFERENCE: 99-82
; CURRENT APPLICATION NUMBER: US/09/781,080B
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-781-080B-2

Query Match 99.6%; Score 770; DB 10; Length 203;
Best Local Similarity 99.3%; Pred. No. 1.3e-30;
Matches 147; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGGILHLELLVAVGPDVFOAHOEDTERVYLTNLNIGALLRDPSSLGAQFRVHLVKWVIL 60
Db 2 AAGGILHLELLVAVGPDVFOAHOEDTERVYLTNLNIGALLRDPSSLGAQFRVHLVKWVIL 61
QY 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQVRG 120
Db 62 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQVRG 121
QY 121 VTOLGGACSPWTSCLITDGTGFDLGVTI 148
Db 122 VTOLGGACSPWTSCLITDGTGFDLGVTI 149

RESULT 6

US-09-781-080B-11
; Sequence 11, Application US/09781080B
; Patent No. US20020142439A1
; GENERAL INFORMATION:
; APPLICANT: HOLLOWAY, JIM
; APPLICANT: SHEPPARD, PAUL
; APPLICANT: YAMAMOTO, GAYLE
; TITLE OF INVENTION: Anti-Angiogenic Intestinal Peptides,
; TITLE OF INVENTION: zdn125
; FILE REFERENCE: 99-82
; CURRENT APPLICATION NUMBER: US/09/781,080B
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 1120
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1120)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-781-080B-11

Query Match 99.6%; Score 770; DB 10; Length 1120;
Best Local Similarity 99.3%; Pred. No. 6.7e-29;
Matches 147; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLVAVGPDVFOAHOEDTERVYLTNLNIGALLRDPSSLGAQFRVHLVKWVIL 60
Db 105 AAGGILHLELLVAVGPDVFOAHOEDTERVYLTNLNIGALLRDPSSLGAQFRVHLVKWVIL 164
QY 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQVRG 120
Db 165 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQVRG 224
QY 121 VTOLGGACSPWTSCLITDGTGFDLGVTI 148
Db 225 VTOLGGACSPWTSCLITDGTGFDLGVTI 252

RESULT 7

US-09-833-328-15
; Sequence 15, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Scheifflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Randolph
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Gerritsen, Helena E.
; TITLE OF INVENTION: Compositing a Polypeptide Chain with the Amino Acid Sequence
; FILE REFERENCE: 247,00CIP
; CURRENT APPLICATION NUMBER: US/09/833,328
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 136
; TYPE: PRT
; ORGANISM: human
US-09-833-328-15

Query Match 92.6%; Score 716; DB 10; Length 136;
Best Local Similarity 100.0%; Pred. No. 2.4e-28;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 AVGPDVFOAHOEDTERVYLTNLNIGALLRDPSSLGAQFRVHLVKWVILTEPEGAPNITAN 72
Db 1 AVGPDVFOAHOEDTERVYLTNLNIGALLRDPSSLGAQFRVHLVKWVILTEPEGAPNITAN 60
QY 73 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQVRGVTOLGGACSPW 132
Db 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQVRGVTOLGGACSPW 120
QY 133 SCLITDGTGFDLGVTI 148
Db 121 SCLITDGTGFDLGVTI 136

RESULT 8

US-09-833-328-2
; Sequence 2, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Scheifflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Randolph
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus

APPLICANT: Furlan, Miha
APPLICANT: Turecek, Peter
APPLICANT: Gerritsen, Helena E.
TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
FILE REFERENCE: 247.00CIP
CURRENT APPLICATION NUMBER: US/09/833,328
CURRENT FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 09/721,254
PRIOR FILING DATE: 2000-11-22
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 133
TYPE: PRT
ORGANISM: human
US-09-833-328-2

Query Match 90.8%; Score 702; DB 10; Length 133;
Best Local Similarity 100.0%; Pred. No. 1.1e-27;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 PDVFOAQHEDTERVLTNLNIGALLRPSLGAQFRVHLVKMVLTEPEGAPNITANLTS 75
DB 1 PDVFOAQHEDTERVLTNLNIGALLRPSLGAQFRVHLVKMVLTEPEGAPNITANLTS 60
QY 76 SLLSVCWGSQTINPEDDTPGHADLVLYITRFDELDPGNROVRGVTOLGGACSPWACL 135
DB 61 SLLSVCWGSQTINPEDDTPGHADLVLYITRFDELDPGNROVRGVTOLGGACSPWACL 120
QY 136 ITEDTGFDLGVTI 148
DB 121 ITEDTGFDLGVTI 133

RESULT 9
US-09-759-508B-2
Sequence 2, Application US/09759508B
Publication No. US20020182599A1
GENERAL INFORMATION:
APPLICANT: Fishman, Mark C.
TITLE OF INVENTION: Methods for Diagnosing and Treating Heart Disease
FILE REFERENCE: 00786/381002
CURRENT APPLICATION NUMBER: US/09/759,508B
CURRENT FILING DATE: 2001-01-12
PRIOR APPLICATION NUMBER: US 60/175,787
PRIOR FILING DATE: 2000-01-12
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 26926
TYPE: PRT
ORGANISM: Homo sapiens
US-09-759-508B-2

Query Match 35.2%; Score 272; DB 9; Length 26926;
Best Local Similarity 0.4%; Pred. No. 0.4;
Matches 101; Conservative 16; Mismatches 23; Indels 24887; Gaps 28;
QY 9 ELL----- 11
DB 190 ELLVOGEEVPKTKTIVSTAQISESROTRIEKKIEAHFDARSIAIVEMVIDGAAGQOL 249
QY 12 ----- 11
DB 250 PHKTPPIPPKPSPTPPSIAAKAQLARQOSPPIRHSPPVRHVRAPTSPSVRSVP 309
QY 12 -----VAVGPDV----- 18
DB 310 AARISTSPVRSPLLMKRTQASTVATGPEVPPPKQEGVASSSEAMRETTLTSTQ 369
QY 19 ----- 18

Db 370 IRTEERMEGRYGVQEOQVTISGAAGAAASVSASAYAAEAVATGAKEVKODADKSAAVATV 429
QY 19 ----- 18
Db 430 VAAVDMARVREPVISAVEQTAORTTTTAVHIQPAQEOVRKEAKTAVTKVVVAADKAKEQ 489
QY 19 ----- 18
Db 490 ELKSRTKEIITTKQEQMHVTHQIRKETKTFVPKVVISAAKAKEQETRISSEITKKQKQ 549
QY 19 ----- 18
Db 550 VTOEAIMKETRKTVPVKVIVATPKVKEODLVSRGREGITTKREBOVOITOEKMRKEAKTA 609
QY 19 ----- 18
Db 610 LSTIAVATAKAKEQETILTRTETMATRBOEQIVTHGKVDVGKAEAVATVVAADOARVR 669
QY 19 ----- 18
Db 670 EPREPGHLEESYAQOTTLEYGYKERISAAKVAEPPORPASEPHVVPKAVKPRVIOAPSET 729
QY 19 ----- 18
Db 730 HIKTDOKGHMISSQIKKTTDLTTERLVHVKRPTASPHFTVSKISVPKTEHGYEASIA 789
QY 19 ----- 18
Db 790 GSAIATLQKELSATSSAQKITKSVKAPTVPKSETRVRAEPTPLQPPFPADTPTDYKSEAG 849
QY 19 ----- 18
Db 850 VEYKEVGSITGTTVREERFEVLHGREAKVTETARVPAPVEIPVTPPTLVSLKXNVTVI 909
QY 19 -----FQ-----AQEDTERYV- 30
Db 910 EGESVTLECHISGYSPSTVWYREDYQIESSIDFOITFOSGIARLMIREAFEDSGRFTC 969
QY 31 ----- 30
Db 970 SAVNEAGTVSTCYLAVQVSEBEFETKETTAVTEKFTTEBKRFRVESRDVMTDTSLTEQNG 1029
QY 31 ----- 30
Db 1030 PGEPAAPYFITKPVVQKLVEGSGVVGCGVGNPKPHVYKKGSLTGTGYRYKVSYNKQ 1089
QY 31 ----- 30
Db 1090 TGECKLVISMTFADDAGEYTVVRNKHGETSASASLLEADYELLMKSQOEMLYOTQVTA 1149
QY 31 ----- 30
Db 1150 FVQPEVGETAPGFVYSEYEKEQEALIRKXMAKDTVVVTVVEDQEPHISSEERLI 1209
QY 31 ----- 30
Db 1210 KEIERYIKTTLLEELLEDGEKMAVDISEASESGDLRIKXRYILEGMVGFPHCKMS 1269
QY 31 ----- 30
Db 1270 GYPLKIAWKDGKRIKHGERVQMDFLQDGRASLRIPVLPDEGIYTAFAFNKIGNAIC 1329
QY 31 ----- 30
Db 1330 SGKLYVEPAAPLGAPTYIPTLEPVSRIRSLSPRSVSRSPIRMSPARMSPARMSPA 1389
QY 31 ----- 30
Db 1390 RMSPCORLEETDESOLERLYKPVFVLKVPFKCLEGOTARFDLVKVGCRPMPTFWHDQO 1449
QY 31 ----- 30
Db 1450 QIVNDYTHKVVIKEDGTOSLIIVPATPSDSGEMTVAQNAGRSSISVILTVEAVEHQV 1509

```
QY 31 -----LTNLNI----- 36
Db 1510 PMFVEKLNKVNIEKSGSRLEMKVRATGNPNPDI VMLKNSDIIVPHKYPKIRIEGTGGEAAL 1569
QY 37 ----- 36
Db 1570 KIDSTVSODSAWYTATAINKAGRDTRCKVNVVEFAEPEPERKLIIPRGTYRAKEIAAP 1629
QY 37 ----- 36
Db 1630 ELEPLHLRYGOBOWEBGLDYDKEKQKPFKKLTSLRLKRGPAHFECLRLPISDPTWV 1689
QY 37 ----- 36
Db 1690 VEWLHDGKPLEAANRLRMINEFCYCSLDYGVAYSRDSGIITCRATNKYGTDHTSATLIVK 1749
QY 37 -----GA----- 38
Db 1750 DEKSLVEESOLPEGRKGLQRIEELERMAHEGALTGVTTDQKEKQKPDIVLYPEPVRVLEG 1809
QY 39 ----- 38
Db 1810 ETARFRVRTGYPPQKKNVYLNQGLIRKSRFRVYDGIHYLDIVDCKSYDGTGEVKVTAE 1869
QY 39 ----- 38
Db 1870 NPEGVIEHKVLEIQOEDFRSVLRRAPRPEFHVHEPGKLOFEVQKVDRPVDTTETKE 1929
QY 39 -----ELLR--- 42
Db 1930 VVKLKRAERITHEKVPSESELRSKFRRTTEGYEAITAVELKSRKSDSEYELLRKTK 1989
QY 43 ----- 42
Db 1990 DELLHWTKELTEEEKALAECKITIPFKPKDKIELSPSWEAPKIPERIQSQTVOGSDA 2049
QY 43 -----DPS----- 45
Db 2050 HFRVRVVGKPDCECEYKNGVKIERSDRIYVWYWPEDNVCELVIRDVTAEDSASIMVKAIN 2109
QY 46 ----- 45
Db 2110 IAGETSSHAFLVQAKLITFTOELQDVVAKEDTWTATPECETSEPFVKVWKYKDGMEVH 2169
QY 46 ----- 45
Db 2170 EGDKYRMHSDRKVHFLSILTIDTSDAEDYSCVLVEDENVKTTAKLIVEGAVVEFVKELQD 2229
QY 46 ----- 45
Db 2230 IEVPESYSGELECIUSPENIEGKWYHNDVELKSNKGYTITSRGRQNLTVDKVTKEHQE 2289
QY 46 ----- 45
Db 2290 YSFVIDGKTKTCKLXMKPRPIAILQGLSDQKVEGDIQVLEVKVLSLESVEGYMMKDGQEV 2349
QY 46 ----- 45
Db 2350 QPSDRVHIVIDKQSHMLLIEDMTKEDAGNYSFTIPALGLSTSGRVSVYSVDVITPLKDVN 2409
QY 46 ----- 45
Db 2410 VIEGTKAVLECKVSPDVTSVKVYLNDEQIKPDDRVQAIKVGTKORLVINRTHASDEGYP 2469
QY 46 ----- 45
Db 2470 KLIIVGRVETNCNLSVEKIKIIRGLRDLTCTETQNVVFEVLSHSGIDVLWNFKDKXIKPS 2529
QY 46 ----- 45
Db 2530 SKYKIEAHGKIYKLTVLNMKDEGKYTFYAGENMTSGKLTVAGGAISKPLTDQTVAESQ 2589
QY 46 ----- 45
Db 2590 EAVFECEVANPDSKGEMLRDGKHLPLTNNIRSESDGHRRLIIAATKLDDIGEYTVKVAT 2649
QY 46 ----- 45
Db 2650 SKTSAKLKVEAVKIKKTLKNLTVTETQDAVFTVELTHPNVKGVMQIKNGVLESNEKYAI 2709
QY 46 -----LGAQFRVHL--VKMV-- 58
Db 2710 SVKGTIYSLRIRKNCALVDESYYGFRGLRGLGASARLHVETVKIIKKPKDVTALENATVAPE 2769
QY 59 ----- 58
Db 2770 VSVSHDTVPVKWPHKSVKIPSDKHRLVSRKVHKLMLQNLSPSDAGEYTAVVGOLECKA 2829
QY 59 ----- 58
Db 2830 KLFVETLHITKMKNIIEVPTKTASFCEVSHFNVPMMKNGVEIEMSEKPIVVQGLK 2889
QY 59 ----- 58
Db 2890 HOLIIMNTSTEDSAEYTFVCGNDQVSATLTVTPIMITSMKLDINAEKDTITFEVTVNYE 2949
QY 59 ----- 58
Db 2950 GISYKWLKNGVEIKSTDKCOMRTKKLTHSLNIRNVHFGDAADYTFVAGKATSTATLYVEA 3009
QY 59 ----- 58
Db 3010 RHIEFRKHIKDIKVLKRAMFECEVSEBPDITVOMKDDOELQITDRIKIOKEYVHRL 3069
QY 59 ----- 58
Db 3070 IPSTRMSDAGKYTVVAGGNVSTAKLFVEGRDVRIRSIKKEVQVIEKQRAVVEFVNEDDV 3129
QY 59 ----- 58
Db 3130 DAHWYKDGIEINFQVERHKYVVERRIHRMPFISETRQSDAGEYTFVAGNRSSVTLVYNA 3189
QY 59 ----- 58
Db 3190 PEPPQVLOELQVPTVQSGKPARFCAMISGRPOPKISWYKEBOLLSTGPKCKFLHDGQEY 3249
QY 59 ----- 58
Db 3250 LLLIEAFPEDAAYTCEAKNDYGVATTASUSVEVPEVSPQEMPVTPPAIITPLQDTV 3309
QY 59 ----- 58
Db 3310 TSEGOPARFQCRVSGTDLKVSHYSKDKKIKPSRPFRTMQPEDTYOLEIAEAYPEDEGTYT 3369
QY 59 ----- 58
Db 3370 FVANNAVQVSSSTANLSLEAPESILHERIEOEIEMKEPSSSPLSABEGLHSAELQLS 3429
QY 59 ----- 58
Db 3430 KINETLELLSESPVYTKFDSEKGTGPIFIKEVSNADISMGDVATLSVTVIGIPKPKIQ 3489
QY 59 ----- 58
Db 3490 WFNCGVLLTPSADYKFPDGGDDHSLIILFTKLEDEGEYTCMASNDYDKTICSAYLKINSK 3549
QY 59 ----- 58
Db 3550 GEGHKDTESAVAKSLEKLGCPPHFLKELKPIRCAQGLPAIFEYTVVGEPAPTVTFW 3609
QY 59 ----- 58
Db 3610 KENKOLCTSVYYTIIHNPNSGSTFIVNDPOREDSGLYICKAENMLGESTCAAEALLVLED 3669
QY 59 ----- 58
```

Db	3670	TDMTDFCKAKSTPEAPEDFPOTPLKGPVAVELDSQEIATFVKDTILKAALITENQOL	3729	Db	4750	DGKIVVEKGRIVPGVIGLMRALTINDADDTAGTYTVTVENANNLECCSCVKVVEIRD	4809
Qy	59	-----	58	Qy	70	-----	69
Db	3730	SYEHIAKANELSSQLPLGAQELOSILQODKLTPESTREFLCINGSIHFOPLKEPSNLQL	3789	Db	4810	WLVPKPIRDOHVVKPCTAIPACDIANKTBNIKWFKGYDEIPAPBNDKTEILRDGNHLYLKI	4869
Qy	59	-----ILTEPEG- : 	65	Qy	70	-----	69
Db	3790	QIVOSOKTFSKEGILMPEEPETOAVLSDETEKIFPMSAMSIQINSLSVTPEPLKTLAEPEGN	3849	Db	4870	KNAMPEDIAEYAVEIEGKRYPAKLTGEREVELLKPIEDVTIYEKESASFDAEISEADIP	4929
Qy	66	-----	65	Qy	70	-----	69
Db	3850	YPOSSIEPPMHSYLTVSAEVLSEKENTVSDTNRQRTVLOKQEAQSALISLSLAEGHV	3909	Db	4930	GQWKLKGELLRPSPTCEIKAEGGKRFLLHKVKLQDAGEVLYQALNAITAILTVKEIEL	4989
Qy	66	-----	65	Qy	70	-----	69
Db	3910	ESLOSPVMISQVNYEPLVPSEHSCTSGKILLIESANPLENAGODSAVRIEEGSKLRFPL	3969	Db	4990	DFAVPLKDVTPERRQARFECVLTREANVINWSKPDPIIKSDKFDIIADGKXHLVINDS	5049
Qy	66	-----	65	Qy	70	-----	69
Db	3970	ALBEKQVLLKEHSDNVMPDQIIESKREPVAIKVQEVQGRDLLSKESLLSGIPBEQR	4029	Db	5050	QPDDEGVYTAIEVEGKTSARLFTVGIRLKFMSPLEDOTVKEGETATFVCELSEKHMVVM	5109
Qy	66	-----	65	Qy	70	-----	69
Db	4030	LNLKIQICRALQAAVASEQPLFSEWLRIEKEVEAVNITQEPRHIMCMVLYTSAKSVT	4089	Db	5110	FKNDAKLHTRSRTVLISSSEKTHKLEMEVTLDDISQIAQVKELSTALQKVLADPYFT	5169
Qy	66	-----	65	Qy	70	-----	69
Db	4090	EEVTIIIEDVDPQMANLKMELRDALCAIIEEIDILTAEGPRIQOGAKTSLQEBMDSFG	4149	Db	5170	VKLHDKTAVEKOBIEILUCEVSKDVPVKWPKDGEIIVSPKYSIKADGLRILKIKKADLK	5229
Qy	66	-----	65	Qy	70	-----	69
Db	4150	SQKVEPIPEVESKYLISTEVSFNVQSRVKYLDATPVTGVASAVSDEKODESLKP	4209	Db	5230	DKGEYVDCGTDKTKANVTVEARLIEVEKPLYGVEVFGETAHFEIELSEPDVHGOWKLK	5289
Qy	66	-----	65	Qy	70	-----	69
Db	4210	SEEKESSSESGTEEVATVKIOEABGLIKEDGPMIHTPLVDTVSEGDIVHLTTSITNA	4269	Db	5290	GQPLTASPDCEIIEDGKHILILHNCQLGMTGEVSPQAANAKSAANLKVKELPLFITPL	5349
Qy	66	-----	65	Qy	70	-----	69
Db	4270	KEVNWYFENKLVPSDEKFKLODQNTYTLVIDKVNTEHDQGEYVCEALNDSGKTATSACL	4329	Db	5350	SDVKVFEKBEAKFECEVSREPKTFRWLKGTOEITGDDRFELIKDGTKHSHVIKSAAFEDE	5409
Qy	66	-----APNI-----	69	Qy	70	-----	69
Db	4330	TVVXSAAPVIKRIKTEPLEVALGHLAKFTCEIOSAPNVRFOWFKAGREIYESDKGIRSSK	4389	Db	5410	AKYMFEBEDKHTSGKLLIEGIRLKLPLTKDVTAKESAVFTVELSHDNIRVKWFKNDQ	5469
Qy	70	-----	69	Qy	70	-----	69
Db	4390	YISSLEILRTQVDCGEYTCASNEYSGVSCATLTVTVPGGEKKVRKLLPERKPEKBE	4449	Db	5470	RLHTRSVSMQDEGKTHSITFKDLSIDDTSQIRVEAMGMSSEAKLTVLEGDPYFTGKLQD	5529
Qy	70	-----	69	Qy	70	-----	69
Db	4450	VVLKSVLRKRPEEBEPKVEPKKLEKVKPAVPEPPPPKPVVEEVPVTVKERKIPEPTK	4509	Db	5530	YTGVEKDEVILQCEISKADAPVKWFKDGEIKPSKNVITDGGKKRMLILKKALKSDIGQ	5589
Qy	70	-----	69	Qy	70	-----	69
Db	4510	VPEIKPAIPLPAPEPKPKEAEVKTIKPPVPEPEPTIAAPVTVPVVGKAEAKAPKEEA	4569	Db	5590	YTCDCGTDKTSGLDIEDREIKLVRPLHSVEVMEETARFETEISEDDIHANWKLKGEAL	5649
Qy	70	-----	69	Qy	70	-----	69
Db	4570	AKPGPIKGVPKKTPSIEAERRKLPGSGGKPPDEAPFTYQLKAVPLKFKVKEIKOIL	4629	Db	5650	LQTPDCEIKEEGKIHSLVHLNCRLDQGGVDFOAANVKSSAHLRVKPRVIGLLRPLKDVT	5709
Qy	70	-----	69	Qy	70	-----	69
Db	4630	TESEFVGSSAIFECVLSPSTAITTWKDGSNIRSPKHFRTADGDKRKLHIIDVOLSAG	4689	Db	5710	VTAGETATFDCELSYEDIPIVEWYLGKGLBPSKVVPKPRSEKVKHTLTLRDVKLESDAGEVQ	5769
Qy	70	-----	69	Qy	70	-----	69
Db	4690	EYTCVLRGNKEKTSTAKLVVEELFVRVFKTLEEEVTVVKGQPLYLSCELNKRVDVWRK	4749	Db	5770	LTAKDFKTHANLKVKEPPVEFTKPLEDOTVEEGATAVLECEVSRENAKVKWFKNGTEILK	5829
Qy	70	-----	69	Qy	70	-----	69
Db	5830	SKKYEIVADGRVKLVIHDCPTPEDIKTYTCDAKDKFTSCNLNVVPPHVEFLRPLTDQVR	5889	Db	5830	SKKYEIVADGRVKLVIHDCPTPEDIKTYTCDAKDKFTSCNLNVVPPHVEFLRPLTDQVR	5889

Qy	70	-----	69	Qy	77	-----	76
Ds	5890	EKEMARFECELSRENAKVKFKDGAIEIKKKYVDYIISKGAVALVINKCLLDDAEAYSE	5949	Ds	6970	LKLENRVKTIISGEIDVNVIAARPSAPKELKFGDITKDSVHLTWBPPDDDDGGSPLTGTVVEK	7029
Qy	70	-----	69	Qy	77	-----	81
Ds	5950	VRTARTSGMLTVLEBEAVFTKLANIEVSETDTIKLVCEVSKPGAIEVIWYKGBEIIETG	6009	Ds	7030	REVSRTKTKVMDFVTDLEFTVPDLVQGEYLFVKVCARNKCGPGEPAYVDEPNVNSTPAT	7089
Qy	70	-----	69	Qy	82	-----	81
Ds	6010	RYEIIITEGRKRILVIONAHLEDAAGNYNCRLPSSRTDGKVKVHLEAAEFISKQNLLEILEG	6069	Ds	7090	VPDPENVKWRDRTANSIFLTWDPKNDGSGRIKGYIVERCPRGSDKWKVACGEPVAETKM	7149
Qy	70	-----	69	Qy	82	-----	81
Ds	6070	EKAIEFVCSISKESFPQMKRDDKILESGDKYDVIADGKKRVLVVKDATLQDMGTVMVVG	6129	Ds	7150	EVTGLEEGKWYAYRVKTLNROGASKSPRTEBEOQADTQEAPEIFLDVKLLAGLTVKAGT	7209
Qy	70	-----	69	Qy	82	-----	81
Ds	6130	AARAAHLTVIEKLRIVVPLKDRVKEQEVVFNCEVNTGAKAKWFRNEEAIFDSSKYI	6189	Ds	7210	KIELPATVGTGPEPKITWTAKADMILKQDKRITIENVPKSTVTIVDSKRSOTGTVIIIEAV	7269
Qy	70	-----	69	Qy	82	-----	81
Ds	6190	ILQOLVYTLRIRDAHLDDQANYNSLVNHRGENVKSAANLIVEBEDURIVEPLKDIETM	6249	Ds	7270	NVCGRATAVVEVNVLDKPGPPAAFDITDVTNESCLLTWNPPRDDGSGKITNVVVERRATD	7329
Qy	70	-----	69	Qy	82	-----	81
Ds	6250	EKKSVTFWCKVNRNLVTLKWTNGEEVFPDNRVSVYRVKYKHMILTIDKCGFPDEGEYIVT	6309	Ds	7330	SEVWHKLSSTVKDTNFKATKLIIPNKEYIFRVAEENMYGAGEPVQASPIAKYQFDPGPP	7389
Qy	70	-----	69	Qy	82	-----	81
Ds	6310	AGQDSVAELLIIEAPTEFVEHEDQTVTFDDAVFSCQLSREKANVKWYNRNGREIKEG	6369	Ds	7390	TRLEPSDITKDAVTLTWCEPDDGGSPITGYWVERLDPDTPKWRNCNMPVKDITTYRVKG	7449
Qy	70	-----	69	Qy	82	-----	81
Ds	6370	KYFKFGSGIHLRIIKDCRLDDEYACGVEDRKSARLFBVEIIPVEIIRPPQDILEAPG	6429	Ds	7450	LTNKKYRFRVLAENLAGPGKPSKSTEPILIKDIPDPWPFGKPTVKDVGKTSVRLNWK	7509
Qy	70	-----	69	Qy	82	-----	81
Ds	6430	ADVFLAELNKQKVEQVLRNMNVVQGDQKQHMSEGKIHRLQICDIKPRDQGEYRFTAK	6489	Ds	7510	PEHDGAKIESVVIEMLKGTGDEWVRVAEGVPTTOHLLPGLMEGOEYSFRVAVNKAES	7569
Qy	70	-----	69	Qy	82	-----	81
Ds	6490	DKEARAKLEAAAPKIKTADQDLVVDVGKPLTMVVPYDAYPKAAEAWFKENEPLSTKTID	6549	Ds	7570	EPSEPSDVLCREKLYPPSPRPLEVINITKNTADLKWTVPEKOGGSPITNIVEKRDVR	7629
Qy	70	-----	69	Qy	82	-----	81
Ds	6550	TTAEQTSFRILEAKGDKGRYKIVLQNHKGKAEGFINKLVDPGVRNLEVTETFDGEV	6609	Ds	7630	RKGQTVDTTVKDTKCTVTPLETSGLYVFRVAANAIGQSDYTEIEDSVLAKDTFTTGP	7689
Qy	70	-----	69	Qy	82	-----	81
Ds	6610	SLAWEPLTDGSKIIGYVVERRDIKRTWVLATDRAESCEFTVTGLQKGVGYLFRVSA	6669	Ds	7690	PYALAVVDVTKRHVDLKWEPKNDGGRPIORVIEKLERLCTRWVKACKTAGPCNPRVT	7749
Qy	70	-----	69	Qy	82	-----	81
Ds	6670	RNRVGTGEPEVDNPEARSKYDVPCLNVITDVRNFGVSLTWEPEYDGAIEITNVV	6729	Ds	7750	DVIEGTEVOFQVRAEENAGVGHPSPTSEILSIEDPTSPPLDLHLVTDAGRKHIAIAWK	7809
Qy	70	-----	69	Qy	82	-----	81
Ds	6730	IELRDKTSIRMDTAMTVRAEDLSATVTDVVEGQESFRVRAQNRIGVKGPSAATPFVKVA	6789	Ds	7810	PPEKNGGSPIIIGYHVEMCPVGTCKMRVNSRPIKDLKPKBEEGVVPDKEYULRVAVNMI	7869
Qy	70	-----	76	Qy	82	-----	81
Ds	6790	DPIERPSPVNLTSDDQTSVOLKWEPLKDGSGPILGYIIERCEEKDNWIRCNMKLV	6849	Ds	7870	GVSEPSSEISENVAKDPCKPTIDLETHDIIIVIEGEKLSIPVPPRAVPVPTVSMHKGDE	7929
Qy	77	-----	76	Qy	82	-----	81
Ds	6850	PELTYKVTGLEKNKYLVRVSAENKAGVSDPSEILGPLTADDAFVEPTMDLSAFKDGLEV	6909	Ds	7930	VKASDRLTMKNDHISAHLEVPKSVRADAGIYITILENKLGSATASINVKVIGLPGCKDI	7989
Qy	77	-----	76	Qy	82	-----	81
Ds	6910	IVPNPITILVSTGPRPTATWCFGDKVLETGDRVKMKTL SAYAELVISPERSDKGIY	6969	Ds	7990	KASDITKSSCKLTWEPPFEGGTPILVYLERREAGRTYIPVMSGENKLSWTVKDLIPN	8049
Qy	77	-----	76	Qy	82	-----	81

Db	8050	GEYFFRVKAVNKVGGGEYIELKNPVIQAQDPKQPPDPVDEVHNPTAEAMTITWKPLPYD	8109
Qy	82	-----	81
Db	8110	GGSKIMGYIIKIAKEERWKRNEHLVPLITYTAKLEBKEYQFVRAENAAGISEPS	8169
Qy	82	-----	81
Db	8170	RATPPTKAVDPIDAPKVLRLTSLEVVRGDEIALDASISGSPYTIITWIKDENVIIVPEEK	8229
Qy	82	-----	81
Db	8230	KRAAPLVRRKGEVQEEFPVLPLOTLSIDNSKKGESQLRVDRSLRPOHGLYMIKVEND	8289
Qy	82	-----	81
Db	8290	HGIAKAPCTVSLDTPGPPINFVEDIRKTSVLCKWEPLDDGGSEIINVTLEKKDKTKP	8349
Qy	82	-----	81
Db	8350	DSEWIVVTSLRHCKYSVTKLIEGKEYLFRVRAENRFGPGPCVSKPLVAKDPFGPPDAP	8409
Qy	82	-----	81
Db	8410	DKPIVEDVTSNMLVKWNEPKDNGSPILGYWLEKREVNSTHWSRVNKSLLNALKANVDGL	8469
Qy	82	-----	81
Db	8470	LEGLTYVFRVCAENAAGPGKSPSPDKTAHDPISPGPPPIPRVTDTSSTTIELEWEPPA	8529
Qy	82	-----	81
Db	8530	FNGGGEIVGYVDKQLVGTNKNWRCRTEKMIKRVQYTVKEIREGADYKLRVSAVNAAGEP	8589
Qy	82	-----	81
Db	8590	PGETQPVTAEPQEPFAVELDVSKGQIQIMAGKTLRIPAVVTGRPVPTKVTWKEGELD	8649
Qy	82	-----	81
Db	8650	KDRVIDNVGTSELIIKDALRKDHGRVITATNSCGSKFAAARVEFVDPGVLDPV	8709
Qy	82	-----	81
Db	8710	VTNRKMCLNWSDEDDGGSEITGFIERKDAKMTWRQPIETERSKCDITGLLEGOEYK	8769
Qy	82	-----	81
Db	8770	FRVIAKNKFGGPPVEIGPILAVDPLGPPTSPERLTYTERQSTITILDWKEPRNGGSP	8829
Qy	82	-----	81
Db	8830	QGYIIKRRHKDPDFERNKRLCPTSLVENLDEHQMYEFRVKAVNAIGESEPSPLNV	8889
Qy	82	-----	81
Db	8890	VIODEVPPTIKLRLSVRGDTIKVAGEPVHIPADVTGLPMPKIENSKNETVIEKPTDAL	8949
Qy	82	-----	81
Db	8950	QITKEEVSREAKTELSIPKAVREDKGTYYTVTASNRLGVSFRNVHVEYDRPSPPRLAV	9009
Qy	82	-----	81
Db	9010	TDIKAESCYLTWADPLDNGSSEITHYVIDKRDAKRKAEMBEVTNTAVEKRYGIWKLPN	9069
Qy	82	-----	81
Db	9070	QOYEFVRVANKYGISDECKSVKVIQDPYRLPGPGPKVLARTKGSMLVSWTPPLDNG	9129
Qy	82	-----	81
Db	9130	GSPITGYWLEKREEGSPYWSRVSRAPITKVGLKGVEFNVPRLLLEGVKYQFRAMAINAACI	9189
Qy	82	-----GW----- 	83
Db	9190	GPPSPSPDEVAGDPIFPBGPPCPEVKDKTKSSISLWKKPPAKDGGSPKCYIVEMQOE	9249
Qy	84	-----	83
Db	9250	GTTDMKRVNEPKLITTCCEVVPNLKELRKVRFRVKAVNAEGESEPSDTTGEIPATDIOE	9309
Qy	84	-----	83
Db	9310	EPEVFIDIGAQCLVCKAGSOIRIPAVIKGRPTPKSSWEFDGKAKKMDGVHDIPEDAQ	9369
Qy	84	-----	83
Db	9370	LETAENSSVIIPECKRSHTGKYSITAKNKAGOKTANCRVKYMDVPGPKDLKVSIDITRG	9429
Qy	84	-----	83
Db	9430	SCRLSMKMPDDGGDRIKGYVIEKRTIDGKAWTKVNPDCGSTTFVVPDLLEQQYFFVR	9489
Qy	84	-----	83
Db	9490	AENRFGIGPPVETIORTTARDPIYPPDPPIKLIKGLITKNTVHLSWKPKNDDGSPVTHY	9549
Qy	84	-----	83
Db	9550	IVECLAWDPTGKKEAWROCNRDVEELOFTVEDLVEGGEYEFBRVKAVNAAGVKPSATV	9609
Qy	84	-----	83
Db	9610	GPCDCQRPDMPPSIDLKEFMEVEEGTNVNI VAKIKGVPPPTLTWFKAPPKPKDNKEPVL	9669
Qy	84	-----	83
Db	9670	DTHVNKLVDVDTCTLVIPOSRRSDTGLYITITAVNNLGTASKEMRLNLVLRPVPVPIKF	9729
Qy	84	-----	83
Db	9730	ESVSADOMTSLWFPKDDGSGKITNVVIEKREANRKTWVHVSSEPKECTYTIPTKLEGHE	9789
Qy	84	-----SOTIN----- 	88
Db	9790	YVFRIMAQNKYGIGELDSEPETARNLFSVPGAPDKPTVSSVTRNSMTVNWEEPEYDGS	9849
Qy	89	-----	88
Db	9850	PVTGYWLEMDTTSKRWKRVNRDPKAMTLGVSYKVTGLIEGSDYQFVYAINAACVGPA	9909
Qy	89	-----	88
Db	9910	SLPSPDPATARDPIAPGPPFPKVDWTKSSADLENSPPLKDGSGKVTGYIIVEYKEGKEE	9969
Qy	89	-----	88
Db	9970	WEGKDKVEVRGTVVTGLKEGAFYKFRVSAVNIAGIGSPGEVTDVIMKDRLVSPDLQL	10029
Qy	89	-----	88
Db	10030	DASVRDRIVVHAGGVIRIIAYVSGKPPPTVTNMNERTLPQEATIIETTAISSMWIKNCQ	10089
Qy	89	-----PEDD----- 	92
Db	10090	RSHQGVYLLAKNEAGERKKTIIIVDLVDPGPGVTGTFFLAHNLNTESKLTFSPEDDGS	10149
Qy	93	-----	92
Db	10150	PITNYVIEKRESDRRAWTPVTYVTRQNAVQGLIQCKAYFFRIAAENSIGMGPFVETSE	10209
Qy	93	-----	92
Db	10210	ALVIREPITVPERPEDLEVKEVTNTVTLTNPPKYDGGSEIINYVLESRLITGEKPHKV	10269

Qy	93	-----	92	Qy	102	-----	101
Db	10270	TNDNLLSRKYTVKGLKEGDTYEVRSVAVNI	VGCGKPSFCTKPTCKDELAPPTLHLDPD	Db	11350	TAOMNRINKDELIROCAFVRPGLIEGNEYRPR	IKAANIVGEGEPRELAESVIADILHPP
Qy	93	-----	92	Qy	102	-----	101
Db	10330	KLTIRVEAFALTGRYSGPKPKVSWFKDEADV	LEDRTTHIKTTPATLALALEKIKAKRSDS	Db	11410	EVELDVTCDVITVRVGTIRILARVGRPEPDI	TWTKGVLRKREKVDLIQDLPVEL
Qy	93	-----	92	Qy	102	-----	101
Db	10390	GKYCVVVNSTGSRKGFQVNVVDHPGPPVGP	VSFEVTKQYVWISWKPPPLDGGSKITN	Db	11470	QIKEAVRADHGYIISAKNSGHAQGSAINVLD	RPGPCQNLKVTNVTKENCTISWENPL
Qy	93	-----	92	Qy	102	-----	101
Db	10450	YIIEKKEVGKVMMPVTSASAKTCKVSKL	LEGKDYIPRIHAENLYGISDPLVSDSMKAK	Db	11530	DNGSSEITNFIVEYRKPKNQKWSIVASDVT	KRLIKANLLANNEYFRVCAENKVGVP
Qy	93	-----	92	Qy	102	-----	101
Db	10510	DRFRVPADPOPIVTEVTKDSALVTWKNP	HDGKPIITNYILEKRETMKRWARTKDPIH	Db	11590	ETKTPILAINIDRPGEPENLHADKGT	FVYLKWRRPDYDGGSPNLSYHVERRLKGSD
Qy	93	-----	92	Qy	102	-----	101
Db	10570	PYTKFRVPDLLEGQYEFVRSAENEIGIDP	SPSPKPVFAKDPIAKPSPPVNPEADTTC	Db	11650	WERVHKSIKETHYMDRCVENIYEFRVOT	KNEGESDMVKTBEVVVKEDLOKPVLDL
Qy	93	-----	92	Qy	102	-----	101
Db	10630	NSVDLTWOPPRHDGSKILGIYIEYQKVG	DEEWRANHTPESCPETKYKVTGLRGOQYK	Db	11710	LSGVLTVKAGDTIRLEAGVRGKPPFVAV	NTKDKDATDLTRS
Qy	93	-----	97	Qy	102	-----	101
Db	10690	FRVLAVNAAGESDPAHVPEPVLVKDR	LEPPELILDANNAHQHIVKVDTRLISAII	Db	11770	KRSDGKYVVTATNTAGSFVAYATVNVLD	KPGPVRLKI
Qy	98	-----	97	Qy	102	-----	101
Db	10750	FPKVTWKEDRDAPTAKRIDVTPVGS	KLEIRNAAHEDGGIYSLTVENPAGSKTV	Db	11830	EIONYILEKCEKTRMVMSTYSATVLT	RGTTVTRLIEGNEYFRVRAENKI
Qy	98	-----	97	Qy	102	-----	101
Db	10810	LDKPGPRDLEVSEIRKDCSYLTWKEP	LDGGSVITNYVVERDVASQWSPLSATSKK	Db	11890	VIAKTKYDKPGRPDPPEVTKVSKKEEM	TVVWNPPEYDGGKSITGYFLEKKEKH
Qy	98	-----	97	Qy	102	-----	101
Db	10870	SHFAKHLNEGNOYLFRVAENQYGRGP	FVETPKPIKALDPLHPGPKDLHHVDV	Db	11950	KSAIPERRRMKVONLLPDHEYQFRV	KAENEIGIGEPSLPSR
Qy	98	-----	97	Qy	102	-----	101
Db	10930	SLVMNKPDRDGGSPITGYLVEYQEG	TDWIKFKTNTNLECVVTLGQOGKTYR	Db	12010	DTTKHSITLHGKPYDGGAPIIGYV	VEMRPKIADASPDEGWKRCNAAOL
Qy	98	-----	97	Qy	102	-----	101
Db	10990	IVGLCLPDTTPIECQEKLVPPSV	ELDVKLIEGLVVKAGTTVRPPAIRGVP	Db	12070	LDENQOEYFRVCAQONQVIGIRPA	ELKEAIKPKKEILPEIDLDASMRKLV
Qy	98	-----	97	Qy	102	-----	101
Db	11050	DGSEIKTDEHVTETDNFSSVLT	IKNCLRRDTGEQITVSNAAGSKTV	Db	12130	FAIVRGRPAPKVTKVRKVIDNVVR	KGOVLDVTMAFLVIPNSTRDDSGK
Qy	98	-----	97	Qy	102	-----	101
Db	11110	PTGPINILDVTFEHTWISWQPPK	DDGGSPVINYIVEKODTRKDTMGV	Db	12190	KAVFVNVVRLDTRGPVSDLKVSDV	TKTSCHVSWAPPENDGGSOVTHIV
Qy	98	-----	97	Qy	102	-----	101
Db	11170	HLOKCEYVFRVAENKICVGPPLD	STPTVAKHKFSPSPGKPVVTDITEN	Db	12250	TVTPEVKKTSFHVNTLVPGNEYY	FRVTAVNVEYGPVTDVPKPVLASD
Qy	98	-----	97	Qy	102	-----	101
Db	11230	PKSDGGSPIITYMERREVCTKWVR	NKTPIDLKFRTVGLYEGNTYEF	Db	12310	ATEMTKNSATLAWLPPLRDGAKID	GYIIISYREBEQPADRWTEYSV
Qy	98	-----	101	Qy	102	-----	101
Db	11290	KPSPSSDPIKACRPIKPPGPPIN	PKLKDKSRETADLVMTKPLSDG	Db	12370	KKYKFRVAARNAGVSLPREAEG	YBAKEQLPPKILMPEQITIKAGK
				Qy	102	-----	101

Db	12430	HPCTCKWKGDEVTSSH LA VH KADSS I I I I I KDVTRKDSGYSLTAENSSGTD TQ K I V	12489	Db	13510	KGKPAPSVWKKGEDPLATDTRVSVESSAVNTT L I V D C Q K S D A G K Y T I T L K N V A C T K E G	13569
Qy	102	-----	101	Qy	102	-----	101
Db	12490	VMDAPGPQPPFDI S I D A D A C S L S M H I P L E D G S N I T N V I E K C D V S R G D W T A L A S V	12549	Db	13570	T I S I K V G K G I P T G P I K F D E V T A E A M T L K W A P K D D G S E I T N V I L E K R D S V N N K W V T C	13629
Qy	102	-----	101	Qy	102	-----	101
Db	12550	T K T S C R V K L I P G O E Y I F R V R A E N R F G I S E P L T S P K M V A Q P F G V P S E P K N A R V T K N K D	12609	Db	13630	A S A V O K T T F R V T R L H E G M E Y T F R V S A E N K Y G V G E G L K S E P I V A R H P F D V P D A P P P N I V D	13689
Qy	102	-----	101	Qy	102	-----	101
Db	12610	C I F V A W D R P D S D G S P I I G Y L I E R K E R N S L L W K A N D T L V R S T E Y P C A G L V E G L E Y S F R I	12669	Db	13690	V R H D S V S L T W T D P K K T G G S P I T G Y H L E F K E R N S L L W K R A N K T P I R M R D F K V T Q L T E G L E Y	13749
Qy	102	-----	101	Qy	102	-----	101
Db	12670	Y A L N K A G S S P S K P T E Y T A R M P V D P P K P E V I D V T K S T V S L I W A R P K H D G S K I I G Y F V	12729	Db	13750	E F R V M A I N A G V K P S L P S E P V A L D P I D P P K P E V I N I T R N S V T L I W T E P K Y D G H K L T	13809
Qy	102	-----	101	Qy	102	-----	101
Db	12730	E A C K L P G D K W R C N T A P H Q I P O E E Y T A T G L E E K A Q Y O F R A I A R T A V N I S P S E P S D P V T I	12789	Db	13810	G Y I V E K R D L P S K S W M K A N H V N P E C A F T V T D L V E G G K Y E F R I R A K N T A C A I S A P S E S T E T	13869
Qy	102	-----	101	Qy	102	-----	101
Db	12790	L A E N V P P R I D L S V A N K S L L T V K A G T N V C L D A T V F G K P M P T V S W K K D G T L L K P A E G I K W A M	12849	Db	13870	I I C K D E A P T I V L D P T I K D G L T I K A G D T I V L N A I S I L G K P L P K S W S K A G K D I R P S D I T	13929
Qy	102	-----	101	Qy	102	-----	101
Db	12850	Q R N L C T L E L F S V N R K S D G V T I T A E N S S G S K S A T I K L V L D K P G P P A S V K I N K M Y S D R A M	12909	Db	13930	Q I T S T P S S M L T I K Y A T R K D A G E Y T I T A T N P F G T K V E H V K V T V L D V G P P G P V E I S N V S A	13989
Qy	102	-----	101	Qy	102	-----	101
Db	12910	L S W E P P L E D G S E I T N V I D K R E T S R P N W A Q V S A T V P I T S C S V E K L I E G H E Y Q F R I C A E N	12969	Db	13990	E K A T L T W T P L S D G S P I K S Y I L E K R E T S R L L W T V V S E D I Q S C R H V A T K L I O G N E Y I F R V	14049
Qy	102	-----	101	Qy	102	-----	101
Db	12970	K Y G V D P V F T E P A I K A N P Y D P P G R C D P P V I S N I T K D H M T V S W K P P A D D G S P I T G Y L L E K	13029	Db	14050	S A V N H Y G K G P V Q S E P V K M V D R F G P P P K P E V S N V T K N T A T V S W K R P V D D G S E I T G Y	14109
Qy	102	-----	101	Qy	102	-----	101
Db	13030	R E T Q A V N W T K V N R K P I I E R T L K A T G L O E G T E Y E F R V T A I N K A G P K P S D A S A A Y A R D P Q	13089	Db	14110	H V E R R E K S L R W R A I K T P V S D L R C K V T G L O E G S T Y E F R V S A E N R A G I P P S E A S D S V L M	14169
Qy	102	-----	101	Qy	102	-----	101
Db	13090	Y P P A P A P K Y Y D T T R S S V S L S W G K P A Y D G S P I I G Y L V E V K R A D S N W V R C N L P Q N L O K	13149	Db	14170	K D A A Y P P P S P N P H V T D T T K S A S L A W K P H Y D G G L E I T G Y V V E H Q K V G D E A W I K D T T G T	14229
Qy	102	-----	101	Qy	102	-LYITRF- 	107
Db	13150	T R F E V T G L M E D T Q Y Q F R V A V N K I G Y S D P S D V P D K H Y P K D I L I P P E G E H D A D L R K T L I L R	13209	Db	14230	A L R I T Q F V V P D L Q T K E K Y N F R I S A I N D A G V G P A V I P D V E I V E R E M A P D F E L D A E L R R T L	14289
Qy	102	-----	101	Qy	108	-----	107
Db	13210	A G V T M R L Y P V K G R P P K I T S K P N V N L R D R I G L I K S T D F T F L R C E N V N K Y D A G K Y I L	13269	Db	14290	V V R A G L S I R I F V P I K G R P A P E V T W K D N I N L K N R A N I E S T F T L I I P E C N R Y D T G K F V	14349
Qy	102	-----	101	Qy	108	-----	107
Db	13270	T L E N S C G K K E Y T I V V K V L D T G P P I N V T V K E I S K D S A Y V T W E P P I I D G S P I I N V V V Q K R	13329	Db	14350	M T I E N P A G K S G F V N V R V L D T P G P V L N L R P T D I T K D S V T L H W D L P L I D G S R I T N V I V E K	14409
Qy	102	-----	101	Qy	108	-----	107
Db	13330	D A E R K S N S T V T T E C S K T S F R V P N L E E G K S Y F F R V F A E N E Y I G D P G E T R D A V K A S O T P G P	13389	Db	14410	R E A T R K S Y S T A T T K C H K T Y K V T G L S E G C E Y F F R V M A E N E Y G I G E P T E T T E P V K A S E A P S	14469
Qy	102	-----	101	Qy	108	-----	107
Db	13390	V V D L K V R S V S K S C S I G W K P H S D G S R I I G Y V V D F L T E E N K W Q R V M K S L S L Q S A K D L T	13449	Db	14470	P P D S L N I M D I T K T S V S L A W P K P H D G S K I T G Y V I E A Q R K G S D W T H I T T V K G L E C V R N	14529
Qy	102	-----	101	Qy	108	-----	107
Db	13450	E G K E Y T F R V S A E N E N G E T P S E I T V A R D D V V A P D L K G L P D L C Y L A K E N S F R L K I P I	13509	Db	14530	L T E G E Y T F Q V M A V N S A G R S A P R S P V I V K E O T M L P E L D R G I Y Q K L V I A K A G D N I K V E	14589
Qy	102	-----	101	Qy	108	-----	107
				Db	14590	I P V L G R P K P T V T W K G D I L K O T Q R V N F E T T A T S T I L I N E C V R S D S G P Y P L T A R N I V G E	14649

Qy	108	-----	107
Db	14650	VGDVITIQVHDIPGPTGPIKFEVSSDFVTSWDPPENDGGVPISNYVVENRQDSTTW	14709
Qy	108	-----	107
Db	14710	VELATTVIRTYKATRLTTGLEQFRVKAQNRGVGPGITSAMIYANYPKVPGPPTQ	14769
Qy	108	-----	107
Db	14770	VTAVTKDSMTISWHPEPLSDGGSPILGYHVERKERNILWQTVSKALVPGNIFKSSGLTDG	14829
Qy	108	-----	107
Db	14830	IAYEFRVAENMAGSKSPSEPMALALDIPDPGKVPVPLNITRHTVTLKMAKPEVTGGF	14889
Qy	108	-----	107
Db	14890	KITSYIVEKRDLPNRWLKNFNSILENEFTVSGLTDAAYEFRVIAKNAAGAISSPSEP	14949
Qy	108	-----	107
Db	14950	SDAITCRDVEAPKIKVDVKFQDVTILKAGEAFRLADVSGRPPPTMEWSKDGKELEGT	15009
Qy	108	-----	107
Db	15010	KLEIKIADFSTNLVKNKSTRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPEGLAYTE	15069
Qy	108	-----	107
Db	15070	VTSEKCVLSWPPPLDGGAKIDHIVQKRETSRLAWTNVASEVQVTKLVTKLLKGNEYI	15129
Qy	108	-----	107
Db	15130	FRMAVKNYGVGEPELESEPVLANPYGPPDPKPNBEVTITKDSMVVCWGHDPDSGGSEI	15189
Qy	108	-----	107
Db	15190	INVIVERDKAQORWIKCNKKTLDLRKYKVSGLTEGHEVEFRIMAENAAGISAPSTSPF	15249
Qy	108	-----	107
Db	15250	YKACDVFPGPPGNPRVLDTSRSSISIAWKNPIYDGGSEITGYMVEIALPEDEWQIVT	15309
Qy	108	-----	107
Db	15310	PPAGLKATSYTITGLTENQEQYKIRIYAMNSEGLGEPALVPGTPKAEDRMLPPEIELDADL	15369
Qy	108	-----	107
Db	15370	RKVVTIRACCTLRLFPVPIKGRDPPEVKWARDHGESLDKASIESASSYTLIIVGNVNRFD	15429
Qy	108	-----	107
Db	15430	GKYILTVENSGSAFVNVRLDTPGPPQDLKVEVTKTSVTLTWDPELLDGGSKIKNY	15489
Qy	108	-----	107
Db	15490	IVEKRESTRKAYSTVATNCHKTSWKVDQLQEGCSYFRVLAENEYIGILPAETAESVKAS	15549
Qy	108	-----	107
Db	15550	ERPLPGKITLMDVTRNSVLSWEKPEHGGSRILGIYIVEMQTKSGDKWATCATVKVTEA	15609
Qy	108	-----	107
Db	15610	TITGLIQEEYSFRVSAQNEKGISDPRQLSVPIAKDLVIPPFAKLLFNTFTVLAGEDLK	15669
Qy	108	-----	107
Db	15670	VDVPFGRPTPAVTHKONVPLKQTRVNAESTENNSLLTIKDACREDVGHYVVKLTNSA	15729
Qy	108	-----	107
Db	15730	GEAETLNVILDKPGPTGPVKMDEVADTSITLSWGPVKYDGGSSINNYIVEKRDSTT	15789
Qy	108	-----	107
Db	15790	TWQIVSATVARTTIKACRLKTGCEYQFRIAAENRYGKSTYLNSEPTVAQYPPKVPGPPTG	15849
Qy	108	-----	107
Db	15850	PVVTLSRDSMEVQWNEPISDGGSRVIGYHLERKERNILWVKLNKTPIPQTKFKTTGLE	15909
Qy	108	-----	107
Db	15910	EGVEYFRVSAENIVGIGKPSKVSECVARDPCDPPGRPEAIIVTRNSVTLQWKKPTVDG	15969
Qy	108	-----	107
Db	15970	GSKITGYIVEKKELPEGRMMKASFTNIIIDTHFEVTGLVEDHRYEPRVIAARNAAGVFSEPS	16029
Qy	108	-----	107
Db	16030	ESTGNAITARDEVDPRIISMDPKYKDTIVVHAGESFKVDADIYKPIPTIOMIKGDOELSN	16089
Qy	108	-----	107
Db	16090	TARLEIKSTDFATSLSVKDAVRVDSGNVILKAKNVAGERSVTNVNKLDRPGPEGPVVI	16149
Qy	108	-----	107
Db	16150	SGVTAECTLAWKPPLODGGSDIINYIVERRETSLRVWTVVDANVQTLSCVKTKLEGN	16209
Qy	108	-----	107
Db	16210	YTFRIMAVNKYGVGEPELESEPVAKNPFVVPDAPKAPBVTVTTKDSMIWVHERPASDGS	16269
Qy	108	-----	107
Db	16270	EILGVLEKRDKEGIRWTRCHKRLIGELRLRVGTGLIENHDYEFVRSNAAGLSEPPSP	16329
Qy	108	-----	107
Db	16330	AYQACDPIYKPGPPNNPKVIDITRSSVFLSWSKPIYDGGCEIOGYIVEKCDNVGEWTM	16389
Qy	108	-----DL 109	
Db	16390	CTPPTGINKTNIEVEKLEKHEYNFRICAINKAGVGEHADVPGPPIIVEEKLEAPDIDL	16449
Qy	110	EL	111
Db	16450	ELRKIIINIRAGSLRFLVPIKGRPTPEVKMGKVDGEIRDAAIIDVTSSTSLVLDNVNRY	16509
Qy	112	-----	111
Db	16510	DSGKYTLTLENSSGTSFAFVTVRLDTPPPVNLKVTEITKDSVSIWEPPLDGGSKIK	16569
Qy	112	-----	111
Db	16570	NYIVEKREATRKSAAVVTNCHKNWKIDQLQEGCSYFRVTAENEYIGILPAQTADPIK	16629
Qy	112	-----	111
Db	16630	VAEVPQPGKITVDVDTNRSVLSWTKPEHDDGSKIIOYIVEMQAKHSEKSECARVKS	16689
Qy	112	-----	111
Db	16690	QAVITNLTOGBEYLFRVAVNEKGRSDPRSLAVPIAKDLVIEPDVKPFAFSYSVQVQD	16749
Qy	112	-----	111
Db	16750	LKIEVPISGRPKPTITWTKDGLPLUKQTRINVTOSLDTLTLUSIKETHKDDGGQYITVAN	16809
Qy	112	-----	111

Db	16810	VWGKTASIEIVTLDPKPPKPVKFDVSAESITLSWNPPLYTGGCQIINYIVQRD	16869
Qy	112	-----	111
Db	16870	TTVWDVVSATVARTTLKVTKLKTGTEYQFRIFAENRYGQSPALSDPIVAQYPYKEGPP	16929
Qy	112	-----	111
Db	16930	GTPATAISKDSMVIQWHEPVNNGSPVIGYHLERKERNISILMTKVNKTIHDTQPKAQN	16989
Qy	112	-----	111
Db	16990	LEEGIEYFRVYAENIVGVGKASKNSECYVARDCDPPGTPPEPIMVKNRNEITLQWTKPVY	17049
Qy	112	-----	111
Db	17050	DGGSMTICYIVEKRDLPDGRMMKASFNTVETQFTVSLGTEDQRYEFRVIAKNAAGAIK	17109
Qy	112	-----	111
Db	17110	PSDSTGITAKDEVELPRISMDPKFRDTIVVNAGETFRLEADVHGKPLPTIEWLRGDK	17169
Qy	112	-----PDGNR	116
Db	17170	EESARCEIKNTDFKALLIIVKDAIRIDGGQYILRASNVAGSKSFVNKVLDRPPEGPV	17229
Qy	117	QVRGVT-----QLGG-----	126
Db	17230	QVTGVTSEKCSLWSPPLDGGSDISHVYVEKRETSRLAWTVVASEVVTNSLKVTKLEG	17289
Qy	127	-----	126
Db	17290	NEYVFRIMAVNKYGVGPESAPVLMKNPFVLPGPSKLEVTNIAKDSMTVCMNRPSDG	17349
Qy	127	-----	126
Db	17350	GSEIIGYIVEKRDRSIRIWKCNKRITDLRLRVTLGTEDHEYFRVSAENAAVGEPSP	17409
Qy	127	-----ACSP-----	130
Db	17410	ATVYKACDPVFKPGPPTNAHI VDTTKNSITLANGKPIYDGGSEILGVVVEICKADEEW	17469
Qy	131	-----	130
Db	17470	QIVTPOTGLRVTRFEISKLEHQBKYIRVCALNKVGLGEATSVPGTVKPEDKLEAPELDL	17529
Qy	131	-----TWS-----	133
Db	17530	DSELRKGVVRAGGSARHIHPFKGRPMPEITWSREGEFTDKVQIEKGVNVTQLSIDNCD	17589
Qy	134	-----	133
Db	17590	RNDAGKYILKLENSGSKSAFVTVKVLDTPGPPQNLA VKEVKDSAFVLEWPEIIDGAK	17649
Qy	134	-----	133
Db	17650	VKNYVIDKRESTRKAYANVSSKSKTSFKVENLTGAIYYFRVMAENEFVGVPVETVDA	17709
Qy	134	-----	133
Db	17710	VKAAPPPPGKVTLTDSQTSASLWKEKPEHDGSGRVLGVYVEMQPKGTEKWSIVAESK	17769
Qy	134	-----	133
Db	17770	VCNAVVTGLSSQEQYQFRVKA YNEKSKDPRLVGLVGPVIAKDLTIQPSLKLPTNTYSIQAG	17829
Qy	134	-----	133
Db	17830	EDLKEIPVIGRPRNISWVKDGEPLKQTRVNVEBTATSTVLHIKEGNKDDFGKYTVTA	17889
Qy	134	-----	133
Db	17890	TNSAGTATENLSVIVLEKPGPPVGPVRFDEVSADFVVISWEPPAYTGGCOISNYIVEKRD	17949
Qy	134	-----	133
Db	17950	TTTTWHMVSATVARTTIKITKLKTGTEYQFRIFAENRYGKSGAPLDSKAVIVQYPPKEPG	18009
Qy	134	-----	133
Db	18010	PGQTPFTSISKDQMLVQWHEPVNDGGTKIIGYHLEQEKNSILVKLNKTPIDTKFKT	18069
Qy	134	-----	133
Db	18070	TGLEGLEYEYFKVSAENIVGIGKPSKVSCEVFARDPCDPPRPEAIVITRNNVTLKWKPK	18129
Qy	134	-----	133
Db	18130	AYDGSKITGYIVEKXDLPDGRMMKASFNTVLETFVSLVEDQRYEFRVIAARNAAGNF	18189
Qy	134	-----	133
Db	18190	SEPSDSSGAITARDEIDAPNASLDPKYKDVIVVHAGETFVLEADIRGKPIPDVVMSKDGK	18249
Qy	134	-----	133
Db	18250	ELEETAARMEIKSTIQKTLVWKDCIRTDGGQVILKLSNVGGTSGIPITVKVLDRLPGSPE	18309
Qy	134	-----	133
Db	18310	GPLKVTGVTAEKCYLAWNPPLODGGANISHYIIEKRETSRLSWTQVSTEVALNKYKTKL	18369
Qy	134	-----	133
Db	18370	LPONEYIFRMAVNVKYGIGEPLESGPVTACNPYKPPGPPSTPEVSATKDSMVVVTWARPV	18429
Qy	134	-----	133
Db	18430	DGGTEIEGYILEKRDKEGVRWTKCNKKTLDLRLRVTLGLTEGHSYEFRAAENAAVGSE	18489
Qy	134	-----	133
Db	18490	PSEPSVFRACDALYPPGPSNPKVTDTSRSSVSLAWSKPIYDGGAPVKGYVVEVEKAAA	18549
Qy	134	-----	133
Db	18550	DEWTTCTPTGLOGKQFTVTKLKENTENYFRICAINSEGVEGEPATLPGSVVAQERIEPPE	18609
Qy	134	-----	133
Db	18610	IELDADLRKVVLRLASATLRLFVTIKGRPEPEVKWEKAEGILTDRAQIEVTSSTMLVID	18669
Qy	134	-----	133
Db	18670	NVTRFDSGRYNTLENNSGSKTAFVNVRLDSPAPVNLTIREVKKDSVTLSEBPLIDG	18729
Qy	134	-----	133
Db	18730	GAKITNIVEKRETRKAYATITNNCTKTTFRIENTLOEGCSYFPRVLASNEYGIGLPAET	18789
Qy	134	-----	133
Db	18790	TEPVKUSEPPLPGORVTLVDVTRNTATIKWEKPEDSGSKITGYVWEMQTKGSEKWSCTCT	18849
Qy	134	-----	133
Db	18850	QVKTLEATISGLTAGEEYVFRVAANVEKGRSDPRQLGVPVIARDIEIKPSVELPFHFNV	18909
Qy	134	-----	133
Db	18910	KAREQLKIDVPFKRPOATVNRKDGOTLKETTRVNVSSKTVTSLSIKEASKEDVGTYE	18969
Qy	134	-----	133
Db	18970	LCVNSAGSITVPITIIIVLDRPGPPGPIRIDEVSCDSITISWNPPEYDGGCOISNYIVEK	19029

QY	134	-----	133	QY	134	-----	133
Db	19030	KETTSTTHIVSOAVARTSIKIVRLTTGSEYQFRVCAENRYGKSSYSESSAVVAEYFPSP	19089	Db	20110	LLKRETSTAVMTSATVARTMMKVMKLTGTGEYQFRIKAENRFGISDHIDSACVTVKLP	20169
QY	134	-----	133	QY	134	-----	133
Db	19090	PGPPGTPKVHATKSTMLVTWQVPVNDGGSRVIGYHLEYKERSILLWSKANKILIADTQV	19149	Db	20170	YTPPGPPTPWNTVNTRESITVGHHEPVNSGSAVGVHLEMKDRNSILWOKANKLVIRT	20229
QY	134	-----	133	QY	134	-----	133
Db	19150	KVSGLDEGLMYEYRVVAENIAGIGKSKSCEPVPARDCPPGQPEVTNITRKSLSLWS	19209	Db	20230	THFKVTTISAGLIYEYRVVAENAGVGKPSHPSEPVLAIDACEPPRNVRITDISKNSVL	20289
QY	134	-----	133	QY	134	-----	133
Db	19210	KPHYDGAKITGIVIVERRELPGDMLKCHYTNIQETYFEVTELTEQRYEYFRVARNAD	19269	Db	20290	SNOOPAFDGGSKITGYIVERRDLPGRWTKASFNTVETQFTISGLTONSOYEFVRFARN	20349
QY	134	-----	133	QY	134	-----	133
Db	19270	SVSEPESTCIIVKDDVEPRVMDVKFRDVIIVKAGEVLKINADIAGRPLFVISWAKD	19329	Db	20350	AVGSISNPSEVVGPITCIDSYGGPVIDPLEYTEVVKVYRAGTSVKLRAGISCKPAPTIEW	20409
QY	134	-----	133	QY	134	-----	133
Db	19330	GIEIERARTEIISTDNHTLLTVKDCIRDTGQYVLTLLKNVAGTRSVAVNCKVLDKPGPP	19389	Db	20410	YKDDKELOTNALVCVENTTDLASILI KQADRLNSCYELKLRNAMASASATIRVOILDKP	20469
QY	134	-----	133	QY	134	-----	CLIT 137 :
Db	19390	AGPLEINGLTAEKCSLSWGRPOEDGGADIDYHRKXRETSHLAWTICEGELQWTSCKVTK	19449	Db	20470	GPPGPIEFKTVTAEKITLLWRPPADGGAKITHYIVREKRETSRVVMSVMSVSEHLBECIIT	20529
QY	134	-----	133	QY	138	-----	137
Db	19450	LLKGNEYIFRVTGVNKYGVGEPLSVAIKALDPFTVPSPPTSLEITSVTKESMTLWSRP	19509	Db	20530	TTKIIKGNEYIFRVRANKYIGIGEPLSDSVVAKNAFVTPGPPGPIEVTKITKNSMTVVM	20589
QY	134	-----	133	QY	138	-----	137
Db	19510	ESDGGSEISGYIIRERKNSLRVRVKNKPYVDLRVKSTGLREGCEYEVYRVAENAAGLS	19569	Db	20590	SRPIADGSDISGYFLEKRDKSLGWFVKVXKETIRDTRQVKTGLTENSQYQYRVCVNAA	20649
QY	134	-----	133	QY	138	-----	137
Db	19570	LPSETSPLRAEDPVLPPSPKPKIVDSGKTTITIAWKVPLFDGAPITGYTVEYKXSD	19629	Db	20650	GQPFSEPFYKAAADPIDPGPPAKIRIADSTKSSITLGWSKPYVYDGGSAVTGYVWEIR	20709
QY	134	-----	133	QY	138	-----	137
Db	19630	DTDWKTSIQSLRGTEYTIISGLTTGAEYVFRVKSVNKVGASDPSDSDPOIAKEREPEPLF	19689	Db	20710	QGESEEWTTVSTKGEVRTTEYVYVSNLKPQVNYFRVSAVNCAGOGEPITEMNEPVOKDIL	20769
QY	134	-----	133	QY	138	-----	137
Db	19690	DIDSEMRKTLIVKAGASFTMTVPFRGRVPVNLWSKPDTLRTRAYVDTTDSRTSLTIEN	19749	Db	20770	EAPEDLDVALRTSVIAKAGEDVQVLI PFKGRPPPTVTRKDEKNLGDARYSIENTDSS	20829
QY	134	-----	133	QY	138	-----	137
Db	19750	ANRDSGKYTLTIONVLSAASLTVVKVLDTPGPPTNITVQDVTKESAVLSWDVPENDGG	19809	Db	20830	SILLTIQVTRNDTGKVIILTIEINGVGEKSTVSVKVLDTPAACQKLQVKHVSRTVTLLW	20889
QY	134	-----	133	QY	138	-----	137
Db	19810	APKNYHIEKREASKAWSVTNNCNRLSYKVTNLQEGAIYFRVSGENEFVGVGIPAEK	19869	Db	20890	DPPLIDGGSPIINYVIEKRDATKRTWSVSHKCSSTSPKLDLSEKTPFFFRVLAENEIG	20949
QY	134	-----	133	QY	138	-----	137
Db	19870	EGVKITEKPSPEKLGVTISKDSVSLTWLKPEHDGGSRIYVHVVEALEKGQKNWYKCAV	19929	Db	20950	IGEPCEETPEVKAEEVPAPIRDLMSKDKSTKTSVILSWTKPDGGSVITEYVVERKQGE	21009
QY	134	-----	133	QY	138	-----	137
Db	19930	AKSTHHVSLRENSEYFRVFAENQAGLSDPRELLLPVLIXEQLBPPEIDMKNFPPSHTV	19989	Db	21010	QTMHAGISKTCEIEVSQLEQSVLEPRVFAKNEKGLSDPVTIGPITVKELIITPEVDLS	21069
QY	134	-----	133	QY	138	-----	137
Db	19990	YVRAGSNLKVDPISCKPLPKVTLSRDGVPLKATMRPNTEITAENLTINLKESVTADAGR	20049	Db	21070	DIPGAQVTVRIGHNVHLELPYKGPSPISMLKGLPLKESEFVRFSKTENKITLSIKNA	21129
QY	134	-----	133	QY	138	-----	137
Db	20050	YEITAANSSGTTKAFINIVULDRPPTGPVVISDITEESVTLKWEPPKYDGGSQVTNVI	20109	Db	21130	KKEHGKGYTILDNAVACRIAPVITVITLGPSPKGPPIRFBDEIKADSVILSWDVPEDNGG	21189
				QY	138	-----	137

Db	21190	GEITCYSIEKRETSOTNWKWVCSSVARTTFKVPNLVKDAEYQFRVRAENRYGVSQPLVSS	21249
Qy	138	-----	137
Db	21250	IIVAKHOPRIQPPGPKPIYVNVTSQMSLTWDAFYDGGSEVTGFHVEKKERNILKQKV	21309
Qy	138	-----	137
Db	21310	NTSPISGREYRATGLVEGLDYQFRVYAENSAGLSPPDSPKFTLAVSPVDPGPDYDV	21369
Qy	138	-----	137
Db	21370	TRETITLKNPPLRDGSKIYGVYIEKXQGNRWRCNFTDVSECOYTVTGLSPGDRYEF	21429
Qy	138	-----	137
Db	21430	RIIARNAVGTISPPSQSGIIMTRDENVPPIVEFGPEYFDGLIIKSGESLRKALVQGRP	21489
Qy	138	-----	137
Db	21490	VPRVTMFKDGVIEIKRMNMEITNVLGSTSLFVRDATRDHRGVYVTEAKNASGSAKAEIKV	21549
Qy	138	-----	137
Db	21550	KVQDTPGKVGPRIREFTNITGEMTLWMDAPLNDGCAPIITHYIIEKRETSRLAWALIEDKC	21609
Qy	138	-----	137
Db	21610	EAQSYTAIKLINGNEYQFRVSAVNFVGVGRPLDSDPVVAQIQYTVDPAPGIPEPSNITGN	21669
Qy	138	-----	137
Db	21670	SITLTWARPESDGGSEIQOYILERREKSTRWVKVISKRPISETRFKVKTGLEGNEYEFH	21729
Qy	138	-----	137
Db	21730	VMAENAVGVGPASGISRLIKREPVPNPGPPTVVVKVTDTSKTTVLSLEWSKPVFDGMEII	21789
Qy	138	-----	137
Db	21790	GYIIEMCKTLGDWHKVNAEACVKTRYTVTDLQAGEBYKFRVSAINGAGKDSCEVTGTI	21849
Qy	138	-----	137
Db	21850	KAVDRLTAPELDIDANFKQTHVVRAGASIRLFIAVQGRPTPTAVWSKPDNSLSURADIHT	21909
Qy	138	-----	137
Db	21910	TDSFSTLTVCNCRNDACKYTLTVENNNGSKSITFTVKVLDTPGPPGPIFTKDVTRGSAT	21969
Qy	138	-----	137
Db	21970	LMDAPLLDGGARIHHYVVEKREASRSWQVISEKTRQIFKVNDLAEGVPYFRVSAVN	22029
Qy	138	-----	137
Db	22030	EYGVGPEYEMPEPIVATEQAPPRRLDVVDTSKSAVLAWLKPDHDGGSRTGYLLENRQ	22089
Qy	138	-----	137
Db	22090	KGSDLNVEAGHTKQLTFTVERLVEKTEYEFVRVAKNDAGYSEPREAFSSVIIKEPQIEPT	22149
Qy	138	-----	137
Db	22150	ADLTGITNQLITCKAGSPFTIDVPISGRPAKVTWKLEMLRKETDRVSIITTKDRITLT	22209
Qy	138	-----	137
Db	22210	VKDSMRGDSGRYFLTLENTAGVKTFSVTVVVIGRGPVGTPIEVSSVSAESCVLWGEPK	22269
Qy	138	-----	137
Db	22270	DGGGTEITNYIVEKRESGTTAQLVNSSVKRTQIKVTHLTKYMEYSFRVSSSENRFVGSKP	22329
Qy	138	-----	137
Db	22330	LESAPIIAEHPFVPSAPTRPEVYHVSAANAMIRWEEPYHDGSKIIGYVWEKKERNITIL	22389
Qy	138	-----	137
Db	22390	WVKENKVPCLCENYKVTGLVEGLEVQFRYYALNAAQVSKASEASRPIMAQNPVDAPGRPE	22449
Qy	138	-----	137
Db	22450	VTDVTRSTVSLIWSAPAYDGGKVGYYIERKPVSEVGDGRMLKCNYYTIVSONFTVTAL	22509
Qy	138	-----	137
Db	22510	SEGDTYEFRLAKNAAGVISKSESTGPTVTCRDEVAPPKAELOARLHGLDLVIRAGSDLV	22569
Qy	138	-----	137
Db	22570	LDAAVCGKPEPKIIMTKGDKELDCEKVSLOYTKRATAVIKFCDRSDSGKYTLTVKNAS	22629
Qy	138	-----	137
Db	22630	GTKAVSMVMKVLDSPGCGKLTVSRVTOEKTCLAWSLPOEDGAEITHYIVERRETSRLN	22689
Qy	138	-----	137
Db	22690	WVIVEGECPTLSYVWVTRLIKNNEYIFRVRANKYQGPVPESEPIVARNSTFTIPSPGIP	22749
Qy	138	-----	137
Db	22750	EEVGTGKEHIIQWTKPESDGGNEISNLYVDKRESLRWTRVKNKYVYVDTRLKVTSLM	22809
Qy	138	-----	137
Db	22810	EGCDYQFRVTAVNAAGNSEPSPSRNFISCREPSYTPGPPSAPRVVDTTKHSISLAWTKPM	22869
Qy	138	-----	137
Db	22870	YDGGTDIVGYVLEMOEKOTDQWYRVHTNATIRNTEFTVPDLKMGOKYSPRVAANVVKGS	22929
Qy	138	-----	137
Db	22930	EYSESTAEIEPVERIEIPDLLEADDLKKTVTIRAGASLRLMVSVSGRPPPVITWSKQGD	22989
Qy	138	-----	137
Db	22990	LASRAIIDTTESVSLIIVDKVNRDAGKYTIEAENQSGKSATVLVKVYDTPGPCPSVKV	23049
Qy	138	-----	137
Db	23050	KEYSRDSVTITWEIPTIDGGAPINNNYVEKREAAAMRAFKVTVTTKCKSLYRISGLVEGTM	23109
Qy	138	-----	137
Db	23110	HYFRVLPENIYGIPECETSDAVLYSEVPLVPAKLEVVVDTKSTVTLAMEKPLYDGGSL	23169
Qy	138	-----	137
Db	23170	TGYVLEACKAGTERWMKVVTLKPTVLEHTVTSINEGEQYLFRIQAQNEKGVSEPRETVTA	23229
Qy	138	-----	137
Db	23230	VTQDURLVPLTIDLSMPOKTIHVPAGRPVELVPIAGRPPPAASWFFPAGSKLRESERV	23289
Qy	138	-----	141
Db	23290	VETHTKVAKLTIRETTIRDTGTYLLEKNVTGTTSETIKVIIIDKPGPTGPIKIDEIDA	23349
Qy	142	-----	141
Db	23350	TSITISWEPPELDGGAPLSGYVVEORDAHRPGWLPVSESVTRSTFKFTRLTEGNEYVFRV	23409

```
QY 142 ----- 141
Db 23410 AATNREGISYQSEVIECRSSIRIPGPETLQIFDVSRDGMLTWYPEDDGGSQVTCY 23469
QY 142 ----- 141
Db 23470 IVERKEVRADRWVRNVKVPVMTWTRYSRSTGLTEGLEYEHRVTAINARGSGKSPRPKPIVA 23529
QY 142 ----- 141
Db 23530 MDPIAPPGKPNPRVTDTRTSVSLAWSVPEDEGGSKVTGYLIEMQKVDQHEWTKCNTTP 23589
QY 142 ----- 141
Db 23590 TKIREVTLTHLPOGAERYFRVLACNAGGPEPAEVPGTVKVTEMLEYPDYELDERVQEGI 23649
QY 142 ----- 141
Db 23650 FVRQGGVIRLTIPIKGKPPFICKWTKEGQDISKRAMIATSETHTELVIKEADRGDSGYD 23709
QY 142 ----- 141
Db 23710 LVLENKCGKANVYIKRVIGSPNSPEGPLEYDDIQVRSVRVSRPPADGGADILGYILE 23769
QY 142 ----- 141
Db 23770 RREVPKAANYTIDSRVGTSLVVKGLKENVEYHFRVSAENQFGISKPLKSEEPVTPKTPL 23829
QY 142 ----- 141
Db 23830 NPPEPPNPPELVDTKSSVSLWSRPKDDGSRVTGYIERKETSTDKWRHNKTQITT 23899
QY 142 ----- 141
Db 23890 TWYTVGLVDAEYOFRIIAQNDVGLSETSPASEVPVCKDPKPSQGELEILSISKDS 23949
QY 142 ----- 141
Db 23950 VTLOWEKPCDGGKEILGYWVEYRQSDSAWKSNNERIKDKQFTIGGLEATEYEFRVF 24009
QY 142 ----- 141
Db 24010 AENETGLSRPRRTAMSIKTLATSGEAPGIRKEMKDVTTKLGEAAQLSCQIVGRPLPDIKW 24069
QY 142 ----- 141
Db 24070 YRFGKELIOSRYKXWSSDGRHTLTVMTTEEQDEGVYTCIATNEVGEVETSSKLLQATP 24129
QY 142 ----- 141
Db 24130 QPHPGYPLKEYYGAVGSTRLHLVMYIGRPVPAMTFHGCQLLQNSENITIENTHYTHL 24189
QY 142 ----- 141
Db 24190 VMKNVQRKTHAGKYKVLNSVFTGVDAILDVEIODPKPTGPVIEALLKNSAVISWKP 24249
QY 142 ----- 141
Db 24250 PADDGGSWITNYVEKCEKAENQVLSSAISVTTCRIVNLTENAGYVFRVSAQNTFGI 24309
QY 142 ----- 141
Db 24310 SDPLEVSSVVIILKSPFEPKPGPKPTITAVTKDSCVAVKPPASDGGAKIRNYLEKREK 24369
QY 142 ----- 141
Db 24370 KONKWSVTBETIRETFVSVKNLIEGLEVEFRVKCNLGESEWSEISEPTPKSDVPIQ 24429
QY 142 ----- 141
Db 24430 APHFKEELRNLRVQSNATLVCKVTGHPKPIVKWYROCKEIIADGLKYRIOEFKGYHQ 24489
```

```
QY 142 ----- 141
Db 24490 LIIASVTDDATVYQVRATNOGGSVSGTASLEVEVPAKIHLPKLTGEGVAHALRGEVVS 24549
QY 142 ----- 141
Db 24550 IKIPFSGKPDPIITWQKQDLIDNNGHYQVIVTRSFSTSLVPPNGVERKDAGFYVVCANR 24609
QY 142 ----- 141
Db 24610 FGIDOKTVELDVADVPDPPRGVKVSDASRDSVNLTWTPASDGGSKITNYIVEKCATTAE 24669
QY 142 ----- 141
Db 24670 RMLRVQARETRYTVINLFGKTSYQFRVIAENKFGLSKPSBPSEPTITKEDKTRAMNDE 24729
QY 142 ----- 141
Db 24730 EVDRETVSMTKASHSSTKELYEKYMIADLGRGEFGIVHRCVETSSKTYMAKFKVKVG 24789
QY 142 ----- 141
Db 24790 TDQVLVKEEISILNIARHNILHLHESFESMEELVMIFEPISGLDIFERINTSAPELNER 24849
QY 142 ----- FDL ----- 144
||:
Db 24850 EIVSVHVQCEALQFLHSHNIGHDFDIRPENIIYQTRRSSTIKIIEFGOARQLKPCDNFRL 24909
QY 145 ----- 144
Db 24910 LFTAPEYVAPEVHQHVSTATDMWSLGLTVYVLLSGINPFLAETNOQIIENIMNAEYTF 24969
QY 145 ----- 144
Db 24970 DEEAFKEISIEAMDFVDRLLLVKERKSRMTASEALQHPWLKOKIERVSTKVIRTLKHRRY 25029
QY 145 ----- 144
Db 25030 HTLIKDLNMVVSAAIRSCGAIIRSQKGSVAKVKVASIEIGPVSGQIMHVGEGGHVK 25089
QY 145 ----- 144
Db 25090 YVCKIENYDOSTQVTWYFGVRQLENSEKEYIYEDGVAILYVKDITKLDDGTYRCKVVD 25149
QY 145 ----- 144
Db 25150 YGEDSSVAELFVKGVREVYDYCRRTMKKIKRRTDTRLLRPPPEFTLPLYNKTYVGEN 25209
QY 145 ---GVTT 148
|||
Db 25210 VRFGVTI 25216
```

```
RESULT 10
US-918-171A-7
; Sequence 7, Application US/09918171A
; Patent No. US20020110894A1
; GENERAL INFORMATION:
; APPLICANT: Aptec, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/04193
; CURRENT APPLICATION NUMBER: US/09/918,171A
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 09/369,364
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
```

US-09-918-171A-7

Query Match 30.7%; Score 237; DB 10; Length 997;
Best Local Similarity 18.2%; Pred. No. 0.011;
Matches 61; Conservative 24; Mismatches 62; Indels 188; Gaps 6;

QY 1 AAGGILHLEL-----PD-----VFOAQHED-----LVAVG----- 15
DB 53 AGGSFLSYELWPRALRKRDVSVRRDAPAFYELQYRGRLRFLNLTANQHLLAPGVSETRR 112
QY 16 -----PD-----VFOAQHED-----LVAVG----- 25
DB 113 RCGLGRAHRAHTPACHLLGEVQDPELEGGAAISACDGLKGVFQLSNEDYFIEPLDSAP 172
QY 26 -----PD-----VFOAQHED-----LVAVG----- 25
DB 173 ARPGAOPHYVYKQAPERLAORGDSSAPSTCGVQVYPELESRRERWEQQRRLRR 232
QY 26 -----PD-----VFOAQHED-----LVAVG----- 25
DB 233 LHQSVSKEKWCETLVADAKWVEYHGOQVESYVLTIMNNAAGLFDHPSIGNPIHITIV 292
QY 56 KWLITPEGAPNITANLTSLLSVCGWSOTINPEDDTPGHADLVLYITRFDELDPGN 115
DB 293 RLVLLEDEEDLKTTHADNTLKSFKWKQKSIKNGKGDARPLHDDTAILLTRKOL-CAAMN 351
QY 116 R--QVRGVTOLGGACSPWMSCLITETDGFDLGVTI 148
DB 352 RPECETGLSHVAGMCQPHRSCSINEDTGLPLAFTV 386

RESULT 11

US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-6

Query Match 30.4%; Score 235; DB 9; Length 11877;
Best Local Similarity 1.1%; Pred. No. 4.1;
Matches 91; Conservative 21; Mismatches 36; Indels 8329; Gaps 25;

QY 1 AAGG-----PD-----VFOAQHED-----LVAVG----- 4
DB 450 AAGGAQVTDVPADRWAGDFYDPRSAPGRSNRWGGFIEDVDRFDAFFGISPREAEM 509
QY 5 -----PD-----VFOAQHED-----LVAVG----- 4
DB 510 DPOORLALGLWEALERAGIDPSSLTGTRGTFAGAIWDDYATLKRGGAAITPPTVTG 569
QY 5 -----PD-----VFOAQHED-----LVAVG----- 4
DB 570 LHRGIIANRLSYTTLGLGPSMWVDSGSSSLVAVHLACESLRGESEALAGVSLNLP 629
QY 5 -----PD-----VFOAQHED-----LVAVG----- 4
DB 630 DSIICASKFGSLSPDGRAYTTFDARANGYVRGEGGFFVVLKLSRAVADGDPVLAVIRGSA 689

QY 5 -----PD-----VFOAQHED-----LVAVG----- 4
DB 690 VNNGAAQMTTPTDAQAEAVLREAHERRAGTAPADVRYVELHGTGTPTVODPIEAAALGAA 749
QY 5 -----PD-----VFOAQHED-----LVAVG----- 9
DB 750 LGTRPAGOPLLVGVSKTNIHLEGNAGTAGLILKAVLAVGRALPASLNYETNPAPPE 809
QY 10 -----PD-----VFOAQHED-----LVAVG----- 9
DB 810 EELNLRVNTVEYLPWEPEDGQRMVGVSVFGMGTTNAHVLEBAPGCGRGASVVESTVGS 869
QY 10 -----PD-----VFOAQHED-----LVAVG----- 9
DB 870 AVGGVYVPMVSAKSAALDAQIERLAAPASRDRDTGVDAGAVDAGAVARVLGG 929
QY 10 -----PD-----VFOAQHED-----LVAVG----- 17
DB 930 RAQFEHRAVVGSGGPDLLAALAAPEGLVRGVASGVGRVAFVPPCGGTOMAGMGAELLD 989
QY 18 --VFQA-----PD-----VFOAQHED-----LVAVG----- 21
DB 990 SAVFAAAMAECEAALSPPYVDSLEAVVROAPGAPTLEVRVDVVPVTFVAVMSLARVMOHH 1049
QY 22 -----PD-----VFOAQHED-----LVAVG----- 21
DB 1050 GVTPOAVVGHSGEIAAAYVAGALSDDAARVVTLSKSIAAHLAKGKGMLSLAUSEDV 1109
QY 22 -----PD-----VFOAQHED-----LVAVG----- 21
DB 1110 LERLAGFDGLSVAANVGPTATVVGDPVQIEELARACADGVRRVIVPDYASHSRQVEI 1169
QY 22 -----PD-----VFOAQHED-----LVAVG----- 21
DB 1170 IESELAELVAGLSPQAPRVFPFSTLEGAMITEPVLDDGGVYWRNLRHRVGFAPAVETLATD 1229
QY 22 -----PD-----VFOAQHED-----LVAVG----- 21
DB 1230 EGFTHFEVSAHPVLTALPGVTGLATLRRDNGGDRLLVASLAERANGLAVDWSPLLP 1289
QY 22 --HOED-----TER-----PD-----VFOAQHED-----LVAVG----- 28
DB 1290 SATGHHSDLPYAFQTERHNLGELIAPAGEPAVQPAVLRTAEAEPAELDRDEQLRVIL 1349
QY 29 -----PD-----VFOAQHED-----LVAVG----- 28
DB 1350 DKVRAQTAQVLGVATGGQIEVDRTFREAGCTSLTGYDLNRINAAFGVRMAPSMPDFPT 1409
QY 29 -----PD-----VFOAQHED-----LVAVG----- 28
DB 1410 PEALAEQLLLHVHGEAANPAGAEPAVAAAGAVDEPVAIVGMACELPGCVASPEDLWRL 1469
QY 29 -----PD-----VFOAQHED-----LVAVG----- 28
DB 1470 VAGGDAISEFPQDRGMVDEGLYHPDPEHPGTSYVYRQGGFIENVAGDAFFGISPREAL 1529
QY 29 -----PD-----VFOAQHED-----LVAVG----- 32
DB 1530 AMDPQORLLLETSSWEAVEDAGIDPTSLRGRQVGVFTGAMTHEYGPFSLRDGGEGLDGYLLT 1589
QY 33 N-----LNIGAEELR-----PD-----VFOAQHED-----LVAVG----- 42
DB 1590 GNTASVMSGRSVYTLGLEPALVTDTACSSSLVALHLAVQALRKEVDNALAGGVAMPT 1649
QY 43 -----PD-----VFOAQHED-----LVAVG----- 42
DB 1650 PGMFVEFSRQRLAGDRSKAFASADGTSWSEGVGVLLVERLSDARRNGHQLVAVRGS 1709
QY 43 -----PD-----VFOAQHED-----LVAVG----- 42
DB 1710 ALNQDGASNGLTAPNGPSQORVIRRALADARLTTSDDVDVVEAHGTGTRGLGDPTEAQAALIA 1769

QY 43 ----- 42
Db 1770 TYGGRDEOPLRLGLSLKSNIGHTQAAAAGVSGVIKQVAMRHGLPKTLHVDESPDQIDW 1829
QY 43 ----- DFLG- 47
Db 1830 SAGAVELLTEAVDPKQDGLRRAAVSFGISCTNAHVLEAPVVPVVEGASVVEPSVVG 1889
QY 48 ----- AOF 50
Db 1890 SAVGGVTPVWVSAAKSAALDAQIERLAASFASRDRTDDADAGAVDAGAVAHVLADGRAQF 1949
QY 51 ----- 50
Db 1950 EHRAVALGAGADBLVQALADPDGLIRGTASGVGRVAFVFPQGTQWAGMAELLDSSAVF 2009
QY 51 ----- RVHLV- 55
Db 2010 AAMAECEAALSPYVDWLSLEAVVRQAPGAPTLERVDVQVQVTPFVVMVSLARVQHGVTP 2069
QY 56 ----- 55
Db 2070 QAVVGHSGEIAAAYVAGALPLDDAARVVTLRSKSTIAAHLAKGGMLSLALNEDAVLERL 2129
QY 56 ----- 55
Db 2130 SDFGLSVAANGPTATVWSGDPVQIELAQACKADGFRARIIPVDYASHSRQVEIESE 2189
QY 56 ----- 55
Db 2190 LAQVLAGLSFQAPRVPFFSTLEGTWITEPVLDTGYWYRNLHRVGPAPAIETLAVDEGFT 2249
QY 56 ----- 55
Db 2250 HFVEVSAHPVLTMLTPETVTGLTLRREGQGERLVTSIAEAWVNGLPVAMTSLLPATAS 2309
QY 56 ----- 55
Db 2310 RFGLPYAFQABRYMELNTPAALATGDDWRYRIDWKRLPAEGBSERTGLSGRLAVTPED 2369
QY 56 ----- 55
Db 2370 HSAQAAVLTALVDAGAKVEVLTAGADDREALAARLTALTGDTGFTGVVSLDGLVQV 2429
QY 56 ----- KVVILTEPE- 64
Db 2430 AMVQALGDAGIKAPLWSVTQAVSVGRLDTPADPDRAMLWGLGRVVVALEHFERWAGLVDL 2489
QY 65 ----- 64
Db 2490 PAQPDAAALHLVLTALSGATGEDQIAIRTTGLHARRLARAPLHGRRTPRDQPHGTVLIT 2549
QY 65 ----- GAPNITANLTSS- 76
Db 2550 GGTGALGSHAARWMAHGAHEHLLVSRSGEQAPGATQUTAEILTASGARVTIAACDVADPH 2609
QY 77 ----- 76
Db 2610 AMRTLLDAIPETPLTAVVHTAGALDDGIVDTLTAEQVRRRAHRAKAVCASVLDLTRLDD 2669
QY 77 ----- 76
Db 2670 LDAFVFSVSTLGIPOGQNYAPHNAYLDALAAARRATGRSAVSAMGPMWGGGMAAGD 2729
QY 77 ----- LLSVCGWSOTI- 87
Db 2730 GVAERLRNHVPGMDPELALAALESALGRDETAITVADIDWDRFYLAYSGRPOPLVEEL 2789
QY 88 ----- 87
Db 2790 PEVRRIIDATSQGGSSAQGANPLAERLAAAAPGERTEILLGLVRAQAAAVLRMRS 2849
QY 88 ----- 87

Db 2850 PEDVAADRAFKDIGFDSLAVELRNRLTRATGLOLPATLVDPHTPLALVSLRSEFLGD 2909
QY 88 ----- NPED- 91
Db 2910 EETADARRSAALPATVGAGAGAGACTDADDPIAIVAMSCRYPGDIRSPEDLWRMLSEGG 2969
QY 92 ----- DTD- 95
Db 2970 EGITFPFTRDGRWDLGLYADPDALGRAYVREGGELHDAEFDAEFFGVSPREALAMPQ 3029
QY 96 ----- 95
Db 3030 QRMILLTTSWEAFERAGIEPASLRGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPS 3089
QY 96 ----- 95
Db 3090 VASGRYATFGLGEPATTVDTACSSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFV 3149
QY 96 ----- 95
Db 3150 EFSRORALAPGRSKAFSADADAGFGAABGVGLLLVERLSDARRNHPVLAVVRGTAVNOD 3209
QY 96 ----- 95
Db 3210 GASNGLTAPNGPSOQVIRQALADARLAPGDI DAVETHGTGTSGLGDPTEAOGLOQTYCKE 3269
QY 96 ----- 95
Db 3270 RPAERPLAIGSVKSNIGHTQAAAGAAGI IKMVLAMRHGTLKTLHADEPSPHVDWANSGL 3329
QY 96 ----- 95
Db 3330 ALVTEPIDWPAGTQPRRAAVSSFCISGTNAHVLEQAPDAAGEVLGADEVSEVSETVAMA 3389
QY 96 ----- 95
Db 3390 GTACTSEVAEGSEASEAPAPGSPREASLPGHLPHVLSAKDEQSLRQQAALHAWLSEPA 3449
QY 96 ----- 95
Db 3450 DLSDADGPARLDVGYTLATSRFAHRAAVTAADRDGFLDGLATLAOGGTSAHVHLDTA 3509
QY 96 ----- GHADL- 100
Db 3510 RDGTTAFLFTGQSORPGAGRELYDRHPVPARALDEICAHLDGHLLEPLLDVMPAABGSA 3569
QY 101 ----- 100
Db 3570 EAALLDETRYTOCALFALEVALPRLVESMGMRPAALLGHSVGEIAAAHVAGWFLSADAAR 3629
QY 101 ----- 100
Db 3630 LVAARGLMOELPAGGAMLAVQAAEDEIRVWLETEBYAGRLDVAAVNGPEAAVLSGDAD 3689
QY 101 ----- 100
Db 3690 AAREAEYWSGLGRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRFRPSLTVVSNVTGL 3749
QY 101 ----- 100
Db 3750 AAGPDDLCDPEYVWRHVGTVRFLDGVRLRDLGVRTCLELGPDGVLTAADAGLADTPA 3809
QY 101 ----- 100
Db 3810 DSAAGSPVGSAGSADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAGHTGPDWH 3869
QY 101 ----- 100
Db 3870 ANFAGSGAHRVLDLPTYSFRDRYWLDA PAADTAVDTAGLGLGTADHPLLCVAVVSLPDRDG 3929
QY 101 ----- 100

Db	3930	LLLTGRLSLRTHPWADHVLGSLVLPGAAMVELAAHAESA	SAGLRDVRRELTLEPLVLPE	3989
Qy	101	-----	-----	100
Db	3990	HGGVELRVTVGAPAGEPGESAGDARPVSLHSRLADAPAGT	AWSCHATGLLATDRPELP	4049
Qy	101	-----	-----	100
Db	4050	VAPDRAAMPPQGAEEVPLDGLYERLDGNGLAFGLPFG	GLNAVWRYEYEFADIALPATT	4109
Qy	101	-----	-----	100
Db	4110	NATAPATANGGSSAAAAPYGIHPALLDASLHATAVGL	GVDEPELVRVPFHWSGVTVHAAG	4169
Qy	101	-----	-----	100
Db	4170	AAAARVRLASAGTDAVSLSLTDGGRPLVSVRRLTP	VTADQAAASRVGGLMHRVAVRP	4229
Qy	101	-----	-----	100
Db	4230	YALASSGEODPHATSYGPTAVLGKDELKVAALAES	AGVEVGLYPDLAALSQDVAAGAPAP	4289
Qy	101	-----	-----	106
Db	4290	RTVLAPLPAGDGAEGVRGTVAITLLELLOAMLAD	EHLAGTRLLLVTRGAVRDPGSGA	4349
Qy	107	-----	-----	106
Db	4350	DGGEDLSHAAAWGLVRTAQTENPGRFGLLDLADD	ASSYRTLPSVLSAGLRDEPQLALH	4409
Qy	107	-----	-----	106
Db	4410	DGTIRLARLASVRPETGTAAPALAPEGTVLLTGG	TGGLGVARHVVGWVRRLLLVSR	4469
Qy	107	-----	-----	106
Db	4470	RGTDPAGADELVHELEALGADVSAACDVADREAL	TAVALDAIPAEHPLTAVVHTAGVLSD	4529
Qy	107	-----	-----	106
Db	4530	GTLPSMTTDEHVLRPKVDAAPFLDLTSTPAYD	LAAFVNFSSAAAVFGAGQGYAAA	4589
Qy	107	-----	-----	106
Db	4590	NATDALAWRRRAAGLPALSLGWGLWAETSQMT	GELQOALRRMSRAGIGISDAEGIAL	4649
Qy	107	-----	-----	106
Db	4650	LDAALRDDRHPVLLPLRLDAAGLRDAAGNDP	AGIPALFRDVVGARTVRRPSAASATTA	4709
Qy	107	-----	-----	106
Db	4710	GTAGTPTGDAEAETAATLADRAATVDGP	PARQRLLEFVVVGEVAEVLGHARGHIDAER	4769
Qy	107	-----	-----	114
Db	4770	GFLDLGFDLSLAVELNRLNSAGGLAUPATLV	FDPHPSAALASHLDAELPRGASDQDGAG	4829
Qy	115	-----	-----	114
Db	4830	NRNGNENGTTASRSTAETDALLAQLTR	LEGALVLTGLSDAPGSEEVLEHLRSLRSMVTGE	4889
Qy	115	-----	-----	114
Db	4890	TGTGTASGAPDGAAGAEADRPWAAAGD	GAGGSEDGAGVPDPFNASABELFGLLDQDPSTD	4949
Qy	115	-----	-----	114
Db	4950	MSTVNEEKYLDYLRATADLHEARGRLRELEAK	AGEPVAIVGMACRLPGGVASPEDLWRL	5009
Qy	115	-----	-----	114
Db	5010	VAGGEDAISEFPQDRGWDVEGLYDNP	PEATCKSYAREAGFLYEAGEFDADFFGISPREAL	5069
Qy	115	-----	-----	114
Db	5070	AMDPOORLLLEASWEAFEHAGIPAATA	RCTSVGVFTGVMYHYATRLTLDVPEGIGYLTG	5129
Qy	115	-----	-----	114
Db	5130	GNSGSVASGRVAYTGLGEPAVTVDTACSS	SLVALHLAVOALRKGEVDMALAGGVTVMST	5189
Qy	115	-----	-----	120
Db	5190	PSTFEVFSQRLGAPDGRSKSFSTADGT	SWSEGVLLVERLSDARRKGRILAVVRGT	5249
Qy	121	-----	-----	120
Db	5250	AVNQDASSGLTAPNGPSOORVIRRALADAR	LTTSDVDVVEAHGTGTRIGDPDIEAOAVIA	5309
Qy	121	-----	-----	120
Db	5310	TYQGRDGEQPLRLGSLKSNIGHTQAAAG	SGVIMVQAMRHCVLPKTLHVEKPTDOVDW	5369
Qy	121	-----	-----	120
Db	5370	SAGAVELLTEAMDWDPKDGGLRRAAVS	FGVSGTNAHVLEBPAAAEETPASEATPAVE	5429
Qy	121	-----	-----	120
Db	5430	PSVGAGLVPLVSAKTPAALDAQIGRLA	APASQGRTDAAADPGAVARVLAGGRAEF	5489
Qy	121	-----	-----	120
Db	5490	VLTGQDDFAQALTAPEGLIRGTPSDV	GRVAVFVFGQGTQWAGMGAELLDVSKEFAAAMA	5549
Qy	121	-----	-----	120
Db	5550	ECESALSRVYDWSLEAVVRQAPCAP	TLERVDVVPVTFAMVSLAKVQHHGVTPQAVVG	5609
Qy	121	-----	-----	120
Db	5610	HSQGEIAAAVYAGALTDDAARVVTL	RSKSIAAHLAKGGMISLASEATRORIENLHG	5669
Qy	121	-----	-----	120
Db	5670	LSIAAVNGPTATVVSQDPTQIOELAQ	ACEADGVRARIIPVDYASHAHVETIESELAEVL	5729
Qy	121	-----	-----	120
Db	5730	AGLSRTPPEVPFSTLEGAWITEP	VLDTGTWYRNLRHVRGAPAVETLATDEGFTHFIEV	5789
Qy	121	-----	-----	125
Db	5790	SAHPVLTMTLPETVTGLTLRREQ	GOERLVTSLAEAWTNGLTIDWAPVLP	TATGHHPEL 5849
Qy	126	-----	-----	125
Db	5850	PTYAFORRHYWLHDSPAVQGS	VQDSWRYRIDWKRLAVADASERAGLSGRWL	VVPVVEDRSA 5909
Qy	126	-----	-----	125
Db	5910	EAAPVLAALSAGADPVLQDVS	PLGDRORLAATLGEALAAAGGAVDGLS	LAWDESAHP 5969
Qy	126	-----	-----	125
Db	5970	GHPAPTRGTGATLTLVQAL	EDAGVAAPLWCVTHGAVSVGRADHVT	SPAQMVMGMRVA 6029
Qy	126	-----	-----	126
Db	6030	ALEHPERWGLIDLPSDADRAAL	DRMTTVLAGGTGEDQVAVRASGLLAR	RLRVLPAHG 6089
Qy	127	ACSPTW-----	-----	132
Db	6090	TASPMQOAGDGLVTGAEP	PAAAEARRLARDGAGHLLHTT	TPSGSEGAECTSGAEDSG 6149

QY	133	-----	132
Db	6150	LAGLVAELADGATATVVTCDLTDAEAAARLLAGVSDAHLPSAVLHLPPTVDSEPLAATD	6209
QY	133	-----	132
Db	6210	ADALARVVTAKATAALHLDRLLREAAAAGRPVVLFFSSVAIWGGAGOGAYAGTAF	6269
QY	133	-----	132
Db	6270	DALAGQHRAQDPTVTSVAVSPWEGSRVTEGATGERLRLGLRPLAPATALTALDTHLGH	6329
QY	133	-----	132
Db	6330	DTAVTIADVWSSFAPGFTTARPCTLLADLPEARRALDEOQSTTAADTVLSRELGALT	6389
QY	133	-----	132
Db	6390	AEQORRMQELVREHLAVVLNHPSPVAVDTGRAFRDLGFSLTAVELNRKLNATGUALPA	6449
QY	133	-----	132
Db	6450	TLVFDYPTPTLAEFLLAETLGEQAGAGEQLPVDGGVDDEPVAIVGMACRLPGGVASPED	6509
QY	133	-----	132
Db	6510	LMRLVAGGEDAISGFPODRGMDVEGLYDPPDASGRTYCRAGGFLDEAGEFDADFFGISP	6569
QY	133	-----	132
Db	6570	REALAMDPOORLLLETSSWEAVEDAGIDPTSLQQQVGVFAGTNGPHYEPLLRNTAEDLEG	6629
QY	133	-----	132
Db	6630	YVGTGNAASIMSRVSYTLGLECPAVTVDTCSSSLVALHLAVQALRKCEGLALAGVT	6689
QY	133	-----	132
Db	6690	VMSTPTTFVFSRQRLAEDGRSKAFASADGFGPAEGVGMVLLVERLSDARRNGHRLAV	6749
QY	133	-----	132
Db	6750	VRGSVAVQDASNGLTAPNGPSQORVIRRALADARLTTADVDVVEAHGTGTRLGDPIEAQ	6809
QY	133	-----	132
Db	6810	ALIATYQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIMVQAMRHGVLPKTLHVDRP	6869
QY	133	-----	132
Db	6870	QIDWSAGTVELLTEAMDPKOEGLIRAAVSSFGISGTNAHIVLEAPVDEDAPEDEPS	6929
QY	133	-----	132
Db	6930	VGVVWPLVSAKTPAALDAQIGRLAFAFASQGRDADPGAVARVLAGGQAQFEHRAVALG	6989
QY	133	-----	132
Db	6990	TGQDDLAALAAPEGLVRGVASGVGRVAFVFPQGTQWAGMCAELLDVSKFPAAMAECE	7049
QY	133	-----	132
Db	7050	AALAPYVWNSLEAVVRQAPGAPTLERVDVVPVTFVAVMVLAKVMQHGVTPQAVVGH	7109
QY	133	-----	132
Db	7110	GEIAAAYVAGALSDDAARVVTLRKSGTAHLAGQGMLSALSSEAVVERLAGFDGLSV	7169
QY	133	-----	132
Db	7170	AAVNGPTATVSGDPTQIOELAQACEADGVRARIIPVDYASHSAHVETIESELADVL	7229
QY	133	-----	132

Db	7230	SPQTPQVPFFSTLEGAWITEPALDGGYWRNLHRHVRGFAPAVETLATDEGTFHFVYSVAH	7289
QY	133	-----	132
Db	7290	PVLTMALPETVGLTGLRRDNGGQRLTTSLAEWANGLTVDWASLLPTTTTHPDLPTVA	7349
QY	133	-----	137
Db	7350	FOTERYWPQDLSAAGDITSAGLGAHEHPLIGAAVALADSOGCLUTGSLRTHPWLADH	7409
QY	138	-----	137
Db	7410	AVAGTVLLPGTAFVELAFRAGDQVGC DLVEELTLDAPLVLRPRGAVRVQLSVGASDESGR	7469
QY	138	-----	137
Db	7470	RTFGLYAHPEDPAGEAEWTRHATGVLAAARADRTAPVADPEAMPVPPGABPVDVGLYERFA	7529
QY	138	-----	137
Db	7530	ANGYCYGPLFOGVRGVHRRGDEVFADVALPAEVAGABGARFGLHPALLDAAVQAAGARG	7589
QY	138	-----	137
Db	7590	VRGHAAAVLERDLLYAVGATALRVRLAPAGPDTVSVSAADSSGQPVFAADSLTVLPVD	7649
QY	138	-----	137
Db	7650	PAQLAAFSDDPTLDALHLEMTAMDGAAQALPGAVVLGGDADGLAAALRAGGTEVLSFPDL	7709
QY	138	-----	137
Db	7710	TDLVEAVDRGETPAPATVTVACPAAGPDGPEHVREALHGSALMQAMLADERFTDGRLLV	7769
QY	138	-----	137
Db	7770	VTRDAVAARSQDGLRSTGQAAVMGLGRSAQTESPCRFLVLLDAGEARTAGDATAGDGLT	7829
QY	138	-----	137
Db	7830	GDATVGTSGDAALGSALATALGSGEPOLARDGALLVPRLARAAAPAAADGLAAADGLA	7889
QY	138	-----	137
Db	7890	ALPLPAAPALWRLEPGTDGSLSLTAAPGDAETLAPPEPLGPGQVRIAIRATGLNFRDVL	7949
QY	138	-----	137
Db	7950	ALGHYDPALMGTEGAGVWTATGVTHTLAPGDRVMGLLSGAYAPVVVADARTVARMPEG	8009
QY	138	-----	137
Db	8010	WTFAGQASVPVFLTAVYALRDLADVKPGRLLVHSAAGVGMAAVOLARHWGVVHGTA	8069
QY	138	-----	137
Db	8070	SHGKWDALRALGLDDAHIASRSLDPESAFRAASGGAGMDVVLNSLARPEFVDASLRLLGP	8129
QY	138	-----	137
Db	8130	GGRFVEMGKTDRDAERVAADHPGVGYRFPDLGEAGPERIGEMLAEVIALFEDGVLRLHP	8189
QY	138	-----	137
Db	8190	VTTMDVRRARDAFRHVSQARHTGKVVLTPSGLDPEGTVLLTGTGTGALGGIVARHVVGW	8249
QY	138	-----	137
Db	8250	GVRRLLLVSRRTDAPCAGELVHELEALGADVSAACDVADREALTAVLDSIPAHEPLTA	8309
QY	138	-----	137

```
Db 8310 VHTAGVLSGDTLPSTAEDEHVLPRKVDAAFLDELSTPGYDLAAAFVMSAAAVFG 8369
Qy 138 ----- 137
Db 8370 GAGQAYAAANATLDALAWRRRTAGLPALSLGWLWLAETSGMTGGLSDTDRSLRSGAT 8429
Qy 138 ----- 137
Db 8430 PMSSELTLSLLDAAMRRDPALVPIALDVAALRAQQRDGLAPLLSGLTGRSGVGGAPVN 8489
Qy 138 -----ED 139
Db 8490 QRRAAAGGAGEADTDLGRLAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAPRD 8549
Qy 140 TGFDL----- 144
Db 8550 TGFDSLTAVELRNRLNAATGLRLPATLVFDHPTPGELAGHLLDELATAAGSGWAEGTSG 8609
Qy 145 ----- 144
Db 8610 DTASATDQTTAAELDRLEGLASLAPAGRPPELAARLALAAALGDDGDDATDLDE 8669
Qy 145 ----- 144
Db 8670 ASDDLFSFIDKELGDSDFMANNEDKRLDKLVRTAELEQQNTRRLREIEGRTHPEVAIVG 8729
Qy 145 ----- 144
Db 8730 MACRLPGVVASPEDLWQLVAGDGAISEFPQDRGWDVEGLYDPPDASGRTYCRSGGFLH 8789
Qy 145 ----- 144
Db 8790 DAGEFADFFGISPREALAMPQORLSLTTAWEAIESAGIDPTALKGSLGVFVGGWHTG 8849
Qy 145 ----- 144
Db 8850 YTGQTAVQSPLEGLHVSQAALGFLSGRIAYVLGTGDPALTVDTACSSSLVALHLAVQ 8909
Qy 145 -----GVTI 148
Db 8910 ALRKGECMDALAGGVTV 8926

RESULT 12
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-6
```

Query Match 30.4%; Score 235; DB 10; Length 11877;
Best Local Similarity 1.1%; Pred. No. 4.1;
Matches 91; Conservative 21; Mismatches 36; Indels 8329; Gaps 25;

```
Qy 1 AAGG----- 4
Db 450 AAGGQAVTVDPADRNAGDFYDPPDRSAFGRSNRGGFIEDVDRFDAFFGISPRAAEM 509
```

```
Qy 5 ----- 4
Db 510 DPQRLALELGNWELERAGIDPSSLTGTRTGVFAGAIWDYATLKHROGGAATPHTVTC 569
Qy 5 ----- 4
Db 570 LHRGIIANRLSYTLGLRGPMSVVDSQSSSLVAVHLACESLRRESELAAGGVSLNLP 629
Qy 5 ----- 4
Db 630 DSIIGASKFGLSPDGRAYTFDARANGYYRGEGGFVWLKRLSRVADGDPVLAVIRGSA 689
Qy 5 ----- 4
Db 690 VNRGAAQGMTPDQAQAEAVLREHERAGTAPADVRYVELHCTGTPVGDPIEAALGAA 749
Qy 5 -----ILHLE----- 9
Db 750 LGTRPAGQPLLVGSKVTNIGHLEGAAGIAGLIKAVLAVGRGALPASLVNVEYTPNPAIPPE 809
Qy 10 ----- 9
Db 810 ELNLRVNTYLPWEPEHGDQRMVGVSPFGMGNTNAHVLEBAPGCRGASVVESTVGS 869
Qy 10 ----- 9
Db 870 AVCGGVVPMVVSXSAALDAQIERLAAFAASRDRTDGVADGAVDAGAVARVLGG 929
Qy 10 -----LLVAVGPD----- 17
Db 930 RAQFEHRAVVVGSGPDDLAALAAAPGLVGVASGVGRVAFVPPGQGTOWAGMGAEILDS 989
Qy 18 --VFOA----- 21
Db 990 SAVFAAAMAECEAALSPPYVDSLEAVVROAPGAPTLERVVDVQPVTFVWVSLARVQHH 1049
Qy 22 ----- 21
Db 1050 GVTPOAVVGHSGQEIAAAYVAGALSLLDDAARVVTLRSKSIAAHLACKGGMLSLALSEDAV 1109
Qy 22 ----- 21
Db 1110 LERLAGDGLSVAANGPTATVVSGDPVQIEELARACEADGVRAARVIPVDYASHSRQVEI 1169
Qy 22 ----- 21
Db 1170 IESELAEVLAGLSQPAPRVFFSTLEGAWITEPVLDDGGYWRNLRHRVGFAPAVETLATD 1229
Qy 22 ----- 21
Db 1230 EGFTHFEVSAHPVLTMALPGTVTGLATLRRDNGDGLVSLAEAWANGLAVDWSPLLP 1289
Qy 22 ----HOED-----TER----- 28
Db 1290 SATCHHSIDLPTVAFQTERHWLGEIALAPAGEPAVQPAVLTEAEPALDRDEQLRVIL 1349
Qy 29 ----- 28
Db 1350 DKVRAOTAQVLGYATGGQIEVDRTPREAGCTSLTGVDLNRNINAATFGVRMAPSMIFDPPT 1409
Qy 29 ----- 28
Db 1410 PEALAEQLLLHVHGEAANAPAGAPAPVAAAGAVDEPVAIVGMACRLPGSVASPEDLWRL 1469
Qy 29 ----- 28
Db 1470 VAGGDAISEFPQDRGWDVEGLYHPDPEHPTGSYYVRQGGFIENVAGDAAFFGISPREAL 1529
Qy 29 -----YVLT 32
Db 1530 AMPQORLLLETSSWEAVEDAGIDPTSLRGRGVQVFTGMANTHEYGPSLRDGGEGLDGYLLT 1589
```

QY 33 N-----LNIGAEELR----- 42
Db 1590 GNTASVMSGRVSYTLGLEGPALTVDTCSSSLVALHLVAQALRKGEVDMALAGGVAVMPT 1649
QY 43 ----- 42
Db 1650 PGMFVEFSRQGLAGGRSKAFASADGTSWSEGVGLLVERLSDARRNGHQLAVVRGS 1709
QY 43 ----- 42
Db 1710 ALNQDGASNGLTAPNGPSOORVIRRALADARLTTSDDVVVEAHGTCTRLGDPLEAQALLIA 1769
QY 43 ----- 42
Db 1770 TYGCRDDEQRLGLSKLSKNIGHTQAAAGVSGVKNVQAMRHGLLPKTLHVDEPSDQIDW 1829
QY 43 -----DPSLG- 47
Db 1830 SAGAVELLTEAVDMPEKQDGLRRAAVSSFGISGTNAHVLEBA PVVVEGASVVEFSVGG 1889
QY 48 -----AQF 50
Db 1890 SAVGGVTPWVVSASAAALDAQIERLAFAFSRDRTDADAGAVDAGAVAHVLADGDAQF 1949
QY 51 ----- 50
Db 1950 EHRAVALGAGADDLVQALADPDGLIRGTASGVGRVAFVPPGGTQWAGMGAELLDSAVF 2009
QY 51 -----RVHLV----- 55
Db 2010 AAAMAECEAALSPYVDWLSLEAVVROAPGAPTLEERVDDVQPVTFVVMVSLARYMHHGVT 2069
QY 56 ----- 55
Db 2070 QAVCHSOGIEIAAAVYAGALPLDDAARVTLRSKSTIAHLACKGGHLSLALNEDAVLERL 2129
QY 56 ----- 55
Db 2130 SDFDGLSVAANGPTATVVGDPVQIEELAQACKADGFRARIIPVDYASHSRQVEIESE 2189
QY 56 ----- 55
Db 2190 LAQVLAGLSPOAPRVPPFFSTLEGTWITEPVLDTWYRNLRHRVGFAPALETAVDEGT 2249
QY 56 ----- 55
Db 2250 HFVEVSAHPVLTMTLPETVTGLTLRREGQGERLVTSLAEAWVNGLPVAVWTSLLPATAS 2309
QY 56 ----- 55
Db 2310 RPGLPTYAQAERYWLENTPAALATCDDWRYRIDMKRLPAEGBERTGLSGRWLAVTPED 2369
QY 56 ----- 55
Db 2370 HSAQAAAVLTALVDGAKVEVLTAGADDDREALAARLTALTTCGDTGVVSLDLGLVPOV 2429
QY 56 -----KVVILTEPE----- 64
Db 2430 AWWQALGDAGIKAPLWSVTQGAHSVGRLDTPADPDRAMLMGLGRVVVALEHPERWAGLVDL 2489
QY 65 ----- 64
Db 2490 PAQPDAAALHLVTALSGATGEDQIAIRTTGLHARRLABAPLHGRRPTRDQPHGTVLIT 2549
QY 65 -----GAPNITANITSS----- 76
Db 2550 GGTGALGSHAARMAHGAELHLLVSRSGEQAPGATQLTAEALTASGARVTIAACDVADPH 2609
QY 77 ----- 76
Db 2610 AMRTLDAIPAETPTTAVVHTAGALDDGIVDTLTAEQVRRRAHRAKAVGASVLDELTRDL 2669
QY 77 ----- 76

Db 2670 LDAFVLFSSVSTLGIPOGNYAPHNAYLDALARRRATGRSAVSAMGPDGGMGAAGD 2729
QY 77 -----LLSVCCWSOTI----- 87
Db 2730 GVAERLRNHGVPMDPELAALAESALCRDETAITVADIDWDRFYLAYSSRPOPLVEEL 2789
QY 88 ----- 87
Db 2790 PEVRRIIDARDSATSGOGSSAOGANPLAERLAAAAPGERTEILLGLVRAQAAAVLRMRS 2849
QY 88 ----- 87
Db 2850 PEDVAADRAFKDIGFDSLAVELRNRLTRATGLOLPATLVFDHPTPLALVLLSRSEFLGD 2909
QY 88 -----NPED----- 91
Db 2910 EETADARRSALPATVGAGAGAGAGTADDDPIAIVAMSCRYPGDIRSPEDULRMLSEGG 2969
QY 92 -----DTDP----- 95
Db 2970 EGITPFPTRDGDLDGLYDADPDALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMPQ 3029
QY 96 ----- 95
Db 3030 QRMILLTTSWEAFERAGIEBPASLRSGSTGVFICLSYQDYAARVPNAPRGVEGYLLTGSTPS 3089
QY 96 ----- 95
Db 3090 VASGRIAYTFGLEGPATTVDTCSSSLTALHLAVRALRSGBCTMALAGVAMMATPHMFV 3149
QY 96 ----- 95
Db 3150 EFSRQALAPDGRSKAFSADADGFGAABGVGLLVERLSDARRNGHPVLAUVRGTVANQD 3209
QY 96 ----- 95
Db 3210 GASNGLTAPNGPSOORVIRQALADARLAPGDIDAVETHGTGTSLGDPTEAOGLOATYKGE 3269
QY 96 ----- 95
Db 3270 RPAERPLAIGSVKSNIGHTOAAAGAAGIIMVLAHRHGTLPKTLHADEPSPHVDWANSGL 3329
QY 96 ----- 95
Db 3330 ALVTEPIDWPAGTGPRRAAVSSFGISGTNAHVLEQAPDAAGEVLGADEVPEVSETVAMA 3389
QY 96 ----- 95
Db 3390 GTAGTSEVAEGSEASEAPAPGSRASLUPGHLPWLSAKDEQSLRGQAAALHJAMLSEPA 3449
QY 96 ----- 95
Db 3450 DLSADGPARLURDVGYTLATSRTPAFHRAAVTAADRDGFLDGLATLAOGGTSAHVHLDTA 3509
QY 96 -----GHADL----- 100
Db 3510 RDGTATFLTCQGSORPCAGRELYDRHPVFARALDEICAHLDGHLELPDLDVMFAEGSA 3569
QY 101 ----- 100
Db 3570 EAALLDETRYTCALFALEVALFRLVSESGMRPAALLGHSVGEIAAAHVAGVFSLADAAR 3629
QY 101 ----- 100
Db 3630 LVAARGRLMQLPAGGAMLAVOAADEIRVWLETEERYAGRLDVAAVNGPEAAVLGSDAD 3689
QY 101 ----- 100
Db 3690 AAREAEATWSGLGRTRRALRVSHAFHSAHMDGMLDGFRAVLETVFEFRPSLTVVSNVTGL 3749
QY 101 ----- 100

Db	3750	AAGPDDLCOPEYVWRHVHVGTVRFPLDGVVRLRDLGVRTCLELPGDGVLTAMAAADGLADTPA	3809
Qy	101	-----	100
Db	3810	DSAAAGSPVGPAGSPADSAAGALRPRPLLVALLRKRSETETVADALGRAHAHGTGPDWH	3869
Qy	101	-----	100
Db	3870	AWFAGSGHRVDLPTYSPRRORYWLDAPAAADTAVDTAGLGLGTADHPLLGAVVSLPDRDG	3929
Qy	101	-----	100
Db	3930	LLLTGRLSLRTHPWLADHAVLGSLVLLPGCAAMVELAAHAESAAGLRDVRRELTLLPLEVLPE	3989
Qy	101	-----	100
Db	3990	HGGVELRVTVGAPAGEGESAGDGARPVSLHSLRLADAPAGTAWSCHATGLLATDRPELP	4049
Qy	101	-----	100
Db	4050	VAPDRAAMWPPQGAEEVPLDGLYERLDONGLAFCPLFQGLNAVWRYEGEVFADIALPATT	4109
Qy	101	-----	100
Db	4110	NATAPATANGGSAAPYGIHPALLDASLHAIAVGGLVDEPELVRVPFHWSGVTVHAAG	4169
Qy	101	-----	100
Db	4170	AAAARVLASAGTDAVSLSLTDGGRPLVSVRLTLRPVTADQAAASRVGGLMHRVAMRP	4229
Qy	101	-----	100
Db	4230	YALASSEQDPHATSYGPTAVLGKDELKVAALLESAGVEVGLYPDLAALSQDVAAAGAPAP	4289
Qy	101	-----VLYITR-----	106
Db	4290	RTVLAPLPAGPADGGAEGVGTAVRTLELLQWLADHLAGTRLLLVTRGAVRDPEGSGA	4349
Qy	107	-----	106
Db	4350	DGGEDLSHAAAGWLVRTAQTNPFRCGLLDLADDAASSYRTLPSVLSDAGLRDEPQALH	4409
Qy	107	-----	106
Db	4410	DGTIRLARLASVRPETGTAAPALAPEGTVLLTGGTGLGGLVARHVVGWVRLLLVSR	4469
Qy	107	-----	106
Db	4470	RGTDAPGADELVHELEALGADVSAACDVADREALTAVLDIAIPAEHPLTAVVHTAGVLS	4529
Qy	107	-----	106
Db	4530	GTLPSMTTDEVHLRPKVDAFLDLDELSTPAYDLAAFMFSSAAAVFGGQGYAAA	4589
Qy	107	-----	106
Db	4590	NATLDALAWRRRAAGLPALSLGWLWAEVSGMTGELCQADLRMSRAGIGISDAEGIAL	4649
Qy	107	-----	106
Db	4650	LDAALRDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVGARTVRRPASAASATTA	4709
Qy	107	-----	106
Db	4710	GTACTPGTADGAETAATVTLADRAATVDGPQRQLLLEFVVGEVAEVLGHARGHRIDAER	4769
Qy	107	-----FDLELPDG-----	114
Db	4770	GFLDGLFDSLTAVELNRNLNSAGGLALPATLVFDHPSPALASHLDLPRGASDQDGAG	4829
Qy	115	-----	114
Db	4830	NRNGNNGTTASRSTAETDALLAQLTRLEGALVLTGLSDAPGSEEVLEHLRSLRSMVTGE	4889
Qy	115	-----	114
Db	4890	TGTGTASGAPDGAGGAEDRPWAAGDAGCGGSEDGAGVPDFMNASAEELFGLLDDQDPSD	4949
Qy	115	-----	114
Db	4950	MSTVNEEKYLDVLRATADLHEARGRLRELEAKAGEPVAIVGMACFLPGGVASPEDLWRL	5009
Qy	115	-----	114
Db	5010	VAGGEDAISEFPQDRGWDVEGLYDPNPEATKSYAREAGFLYEAGEFDADFFQISPREAL	5069
Qy	115	-----	114
Db	5070	AMDPOORLLLEASWEAFEHAGIPAAATARCTSVGVFTGVNMYHDYATRLTLDVPEGIEYLG	5129
Qy	115	-----	114
Db	5130	GNSGSVASGRVAYTLGLEGPVTVDTACSSSLVALHLVAOLRKGEVDMALAGGVTVMST	5189
Qy	115	-----NRQ-----VRG- 120	
Db	5190	PSTFVEFSRQRLAPDGRSKSPSTADGTWSSEGVLLVERLSDARRKGRILAVVRGT	5249
Qy	121	-----	120
Db	5250	AVNQDASSGLTAPNGPQSOORVIRRALADARLTTSDDVDVVEAHGTGTRGLDPIEAQAVTA	5309
Qy	121	-----	120
Db	5310	TYQGRDGEQPLRGLSKSNIGHTOAAAGVSGVIKWQVQMRHGVLPKTLHVEKPTDQVDM	5369
Qy	121	-----	120
Db	5370	SAGAVELLTEAMDMWPKDGLRRAAVSFGVGTNAHVLEAPAAEETPASEATPAVE	5429
Qy	121	-----	120
Db	5430	PSVGAGLVPWLVSATKTPAALDAQIGRLAAFAAQRTDAADPGAVARVLAGGRABEHRV	5489
Qy	121	-----	120
Db	5490	VILGTQDDFAQALTAPEGLIRCTPSDVGRVAFVFPQOGTOWAGMGAELLDVSKFEAANA	5549
Qy	121	-----	120
Db	5550	ECESALLSRYVDWNSLEAVVRQAPGAPTLERVDVVPVTFVAVMVSLAKVMQHGVTPQAVVG	5609
Qy	121	-----	120
Db	5610	HSOGEIAAYVAGALTDDAARVTVLRKSIAHLAGKGMISLALSEBETQRORIENLHG	5669
Qy	121	-----	120
Db	5670	LSIAAVNGPTATVVSQDPTQIOELAQACEADGVRARIIPVDYASHAHVETTESELAEVL	5729
Qy	121	-----	120
Db	5730	AGLSPRTEVPFPFSTLEGAWITEPVLDGTYWYRNLHRVGFAPAVETLATDEGTFHFIEV	5789
Qy	121	-----VTQLG-----	125
Db	5790	SAHPVLTMTLPETVTLGLTLRREGOGBRLVTSABAWTNGLTIDWAPVLPATGHHPEL	5849
Qy	126	-----	125
Db	5850	PTYAFORRHYWLHDSPAVQGSVDQSWRYRIDWKRLAVADASERAGLSGRWLVVVPEDRSA	5909
Qy	126	-----	125
Db	5910	EAAPVLAALSGAGDPVOLDVSPLDGRORLAATLGEALAAAGGAVDGVLSLLAWDESAP	5969

QY	126	-----	125
Db	5970	GHPAFFTRCTGATLTILVQALEDAGVAAPLWCVTHGAVSVGRADHVTSPAQAMVMGMGRVA	6029
QY	126	-----	126
Db	6030	ALEHPERMGLIDLPSDADRAALDRMTTVLAGGTGEDOVAVRASGLLARRLVRASLPAHG	6089
QY	127	ACSPTM-----	132
Db	6090	TASPMQADGTVLVTCABEPAAAEAAARLIARDGAGHLLHHTTPSGSEGAEGTSGAAEDSG	6149
QY	133	-----	132
Db	6150	LAGLVAELADLCATATVVTCDLTDAEAAARLLAGVSDAHPLSAVLHLPTVDSEPLAATD	6209
QY	133	-----	132
Db	6210	ADALARVVTKATAALHLDRLLREAAAAGRPVVLVLFSSVAAIWGGAGGAYAGATFL	6269
QY	133	-----	132
Db	6270	DALAGOHRAHGTVTSVAVSPWEGSRVTEGATGERLRLGLRPLAPATALTALDTALGHG	6329
QY	133	-----	132
Db	6330	DTAVTIADVDSFPAGFTTARP GTLLADLPEARALDEQQSTTAADDTVLSRELCAITG	6389
QY	133	-----	132
Db	6390	AEQORRMQELVREHLAVLNHPSPEAVDTGRAFRDLGFDLSLTAVELNRLKNATGLALPA	6449
QY	133	-----	132
Db	6450	TLVFDYPTERTLAFFLAEILGEQAGAEQLPVDGGVDDEPVAIVGMACRLPGGVASPED	6509
QY	133	-----	132
Db	6510	LWRLVAGGEDAISFPQDRGWDVEGLYDPDPDASGRTYCRAGGFLDEAGEFDADFFGISP	6569
QY	133	-----	132
Db	6570	REALAMDPOORULLETSMWEAVEDAGIDPTSLQQQVGVFAGTNGPHYEP LLRTAEDLEG	6629
QY	133	-----	132
Db	6630	YVGTGNAASIMSGRVSYTLGLEGPVTVDTACSSSLVALHLAVQALRKGEGLAGAGVT	6689
QY	133	-----	132
Db	6690	VNSTPTTTFVEFSRQRGLAEDGRSKAPASADGFGPAEGVGM L LVERLSDARRNGHRVLAV	6749
QY	133	-----	132
Db	6750	VRGSVNDGASNGLTAPNGSQQRVIRRALADARLTADVDVVEAHGTGRLGDP IEAQ	6809
QY	133	-----	132
Db	6810	ALIATYGOGRDTEQLRLGSLKSNIGHTOAAAAGVSGIIKMWQAMRHGVLPKTLHVDRPSD	6869
QY	133	-----	132
Db	6870	QIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGTNAHIVLEEAPVDEDAPEDEPS	6929
QY	133	-----	132
Db	6930	VGVVVPWLVSAKTPAALDAQIGRLAAAFASQGRTDADPGAVARVLAGGRAQFEHRAVALG	6989
QY	133	-----	132
Db	6990	TQODDLAALAAPEGLVRGASGVGRVAFVFPQGTQWAGMGAELLDVSKFEFAAMAEC	7049
QY	133	-----	132

Db	7050	AALAPYVDVSLSEAVVROAPGAPTLLERVDDVQVPTFAVMVSLAKVMQHGHVTPQAVVGHSG	7109
QY	133	-----	132
Db	7110	GEIAAAYVAGALSDDAARVVTLRKSKSIGAHLAGOGGMLSLALSEAAVVERLAGFDGLSV	7169
QY	133	-----	132
Db	7170	AAVNGPTATVVGDPQTOIQELAQACEADGVRARIIPVDYASHSAHVETIESELAADVLAGL	7229
QY	133	-----	132
Db	7230	SPOTPOVBFSTLEGAWITEPALDGGYWYRNLHRVGPAPAVETLATDEGFTHFVSVSAH	7289
QY	133	-----	132
Db	7290	PVLTMALPETVTGLTLLRRDNGGQHLTTS LAEAWANGLTVDWASLLPTTTTHPDLPYIA	7349
QY	133	-----	137
Db	7350	FQTERYWPQDLSAAGDITSAGLGAABHPLLGAAVALADSDGCLLTGSLSLRTHPWLADH	7409
QY	138	-----	137
Db	7410	AVAGTVLLPGTAFVELAFRAGDQVGC DLVBEELTLDAPLVLPRRGAVRVQLSVGASDESGR	7469
QY	138	-----	137
Db	7470	RTFGLYAHPEDAPGEAEWTRHATGVLAAADRTAPADPEAMP PP GAEPVDVDGLYERPA	7529
QY	138	-----	137
Db	7530	ANGYGYGPLFOGVGRVWRGDEVFADVALPAEVAGAEGRFGLHPALLDAAVQAAGACRG	7589
QY	138	-----	137
Db	7590	VRRGHAAAVRLEROLLYAVGATALRLV LAPAGPDTVSVAADSSQCPVFAADSLTVLPVD	7649
QY	138	-----	137
Db	7650	PAQLAAFSDDPTLDALHLLLEWTANGAAQALPGA VVLGGDAGLAAALRAGGTEVLSFPDL	7709
QY	138	-----	137
Db	7710	TDLVEAVDRGETPAPATVLVACPAAGPDGPEHVREALHGS LALMQAWLADERFTDGRLLV	7769
QY	138	-----	137
Db	7770	VTRDAVAARSGDLRSTGQAAVWGLGRSAQTESPGRFVLLDLAGEARTAGMATAGDGUTT	7829
QY	138	-----	137
Db	7830	GDAVGTSGDAALGSLALATALGSGEPQJALURDGALLVPRLARAAAPAAADGLAAADGLA	7889
QY	138	-----	137
Db	7890	ALPLPAAPALWRLEPGTDCSLES LTAAPGDAETLAPEPLGPGQVRIATIRATGLNFRDVL	7949
QY	138	-----	137
Db	7950	ALGMYPDPALMGTEGAGVVTTATPGVTHLAPGCDRVMLLSGAYAPVVVADARTVARMPEG	8009
QY	138	-----	137
Db	8010	WTFAGAGSVVPVFLTAVYALRDLADVKPGERLLVHSAAGVGMVAQVJARHGWVEVHGTA	8069
QY	138	-----	137
Db	8070	SHGKWDALRALGLDDAHIASSRTLD FESA FRAASGGGMDVVV LNSLAREFVDASLRLGCP	8129
QY	138	-----	137

```
Db 8130 GGRFVEMGKTDVDRDAERVAADHPGVGYRAFDLGEAGPERIGEMLAEVIALPEDGVLRHLP 8189
Qy 138 -----
Db 8190 VTTWDRARRDAFRHVSQARHTGKVLVLTMPSGLDPEGVLLTGTGALGGIVARHVVGW 8249
Qy 138 -----
Db 8250 GVRLLLVSRRTDAPGAGELVHEALGADVSAACDVADREALTAVLSDIPAETHPLTA 8309
Qy 138 -----
Db 8310 VHTAGVLSGDTLPSMTAEDVEHVLRPVKVDAFLLDELSTSPGYDLAAFMVFFSAAAVFG 8369
Qy 138 -----
Db 8370 GAGOGAYAAANATLDALAWRRRTAGLPALSLGWLWAEISGMTGGLSDTDRSLRARGAT 8429
Qy 138 -----
Db 8430 PMDSELTLSLLDAAMRDDDPALVPFIALDVAALRAQORDGMLAPLLSGLTRGSRVGGAPVN 8489
Qy 138 -----
Db 8490 QRRAAAGGAGADTLGGRLAAMPDDORVAHLRDLVRTHVATVLGHGTPSRVDLERAFRD 8549
Qy 140 TGFPL-----
Db 8550 TGFPSLTAVELNRNLNAATGURLPATLVFDHPTPELAGHLLDELATAAGGMAEGTSGS 8609
Qy 145 -----
Db 8610 DTASATDRTTAAALAELEDRLEGVLASLAPAGGRPELAARLALAAALGDDGDDATDLDE 8669
Qy 145 -----
Db 8670 ASDDDLFSFIDKELGSDFMANNEDKLDYLRVTAELQNTRELREIEGRTHEPVAIG 8729
Qy 145 -----
Db 8730 MACRLPGVASPEDLWLVAGDGAISEFPQDRGWDVEGLYDPDPDASGRTYCRSGFLH 8789
Qy 145 -----
Db 8790 DAGEFDADFICSPREALAMPDQORLSLTTAWAIESAGIDPTALKSGSLGVFVGWHGTG 8849
Qy 145 -----
Db 8850 YTSQTTAVOSPELEGLVSGAALGFLSGRIAYVLGTGDPALTYDVTACSSSLVALHLAVQ 8909
Qy 145 -----GVTI 148
Db 8910 ALRKGECMDALAGGVTV 8926

RESULT 13
US-09-788-043C-5
; Sequence 5, Application US/09788043C
; Patent No. US20020107361A1
; GENERAL INFORMATION:
; APPLICANT: Heller, Renu
; APPLICANT: Zuo, Fengrong
; APPLICANT: Klonowski, Paul
; TITLE OF INVENTION: No. US20020107361A1el Metalloproteases Having
; TITLE OF INVENTION: Thrombospondin Domains and Nucleic Acid Compositions
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: ROCH-004
; CURRENT APPLICATION NUMBER: US/09/788,043C
; PRIOR FILING DATE: 2001-02-16
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 5
; LENGTH: 1690
; TYPE: PRT
; ORGANISM: human
US-09-788-043C-5

Query Match      29.9%; Score 231; DB 10; Length 1690;
Best Local Similarity 17.9%; Pred. No. 0.071;
Matches 60; Conservative 24; Mismatches 63; Indels 188; Gaps 6;

Qy 1 AAGGILHLEL-----PD-----VFOHQSD-----LVAVG----- 15
Db 57 AGGSFSLSEYELWEPALRKRDVSRRDAPAFYELQYRGRELRFNLTANQHLLAPGVSETRR 116
Qy 16 -----
Db 117 RGGLGRAHRAHTPACHLLGEVODPEGLGAAISACDGLKGVFOLSNEDYFIEPLDSAP 176
Qy 26 -----
Db 177 ARPGHAQPHVYKQAPERLAORGSSAPSTCGVQVYPELEPRERWEORQWRRLRR 236
Qy 26 -----
Db 237 LHQSVSKKQVETLVADAKMVEYHGQPOVESYVLTINMVAGLFHDFPSIGNPIHITIV 296
Qy 56 KMVILTEPEGAPNITANLTSSLLSVCGWSQTNPEDDTPGHADLVLYITRFDLELPDGN 115
Db 297 RLVLLEDEEDLKIITHADNTPKSFCKQKSIKMMKGDAHPLHHDTAILLTRKDL-CATWN 355
Qy 116 R--QVRGVTOLGSGACPTWSCLITEDTGFDLGVTI 148
Db 356 RPECETGLSHVAGMCPHRSCSINEDTGLPLAFTV 390

RESULT 14
US-09-321-987B-2
; Sequence 2, Application US/09321987B
; Patent No. US20020102210A1
; GENERAL INFORMATION:
; APPLICANT: Kimble, Judith E
; APPLICANT: Bleiloch, Robert H
; TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
; FILE REFERENCE: 960296, 95386
; CURRENT APPLICATION NUMBER: US/09/321,987B
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087,170
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/129,023
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2150
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-321-987B-2

Query Match      29.6%; Score 229; DB 10; Length 2150;
Best Local Similarity 3.7%; Pred. No. 0.16;
Matches 61; Conservative 20; Mismatches 65; Indels 1507; Gaps 8;

Qy 3 GGILHL----- 8
Db 5 GGSFHLQPVVAAILLVCLVYALOSGSGTISEFSSDLVFSRAKYSGVPVHHHRWRODA 64
Qy 9 ----- 8
Db 65 GIHVIDSHHIVRRDSYGRGRKRDVTSTDRRRRLOQVARDGCHACHLRLSRDSDAVYVHLH 124
Qy 9 ----- 8
Db 125 RNQIIPDSHNKSVPHFSNSNFAPMVLVLDSEEVGSGMSRTPDCIYRAHVKGVOHSIV 184
```

```

Qy 3 GGTLHL----- 8
      |||
Db 5 GGSFHLQPVAAILLVCLVYALQSGGTISEFSDVLFSSRAKYSQVPVHHSWRQDA 64
Qy 9 ----- 8
Db 65 GIHVDSHHIVRRDSYGRGGRKRDVTSTDRRLQGVARCDCGCHALRLSRDDAVYIVHLH 124
Qy 9 ----- 8
Db 125 RWNQIPDSHNKSPHFSNSNFAPMWLYLDSVEEVGCGMSRTDPDCIYRAHVKGVQHISIV 184
Qy 9 ----- 8
Db 185 NLCDSEDLGYCMALPDSGIHTVEPIISGNGTEHDCASRRHQHLVRKDPWHFKSPDLHNS 244
Qy 9 ----- ELLVAGDPDVFOAQHOEDTERVV 30
      |||
      |||

```



```

Db 245 TSVNETTTTATQWQMEDVIERKARSRRRAANSMDHYVEVLVADTKMYEYHGRSLEDYV 304
Qy 31 LTNLNGAELLRLDPSLQAFRHLVKMVLTEPEGAPNITANLTSSLLSVCGWSQTINPE 90
Db 305 LTLFSTVASIYRHQSRLASINVVVVKLILVKTENAGPRITONAOQTLODFCRWOQYNDP 364
Qy 91 DDTDPGHADLVLYITRFDELPLDGNRQVRGVTOLG----- 125
Db 365 DDSSVOHHDVAILLTRKIDICRSQKCDTLGLAELGTMCDMQSKCAIIEDNGLSAAFTIAH 424
Qy 126 ----- 125
Db 425 ELGHVFSIPHDDERKSTYMPVNVKCKFQSTKFDKTOFQNNFHIMAPLEYNTHPWSWP 484
Qy 126 ----- 125
Db 485 CSAGMLERFLENNRGOTCLOFDPVVERRYEDVFVRDEPGKYDAHQOQKFVFGPASELC 544
Qy 126 ----- 125
Db 545 PYMPTCRLWCATFYGSOMGCRTHMPHADGTPCDESRSMPCHHGACVRLAPESLTKIDG 604
Qy 126 ----- 125
Db 605 QWGDWRSMGECSTRTCGGVQKGLRDCDSPKPRNGKYCVGORERYSCNTOECPWDOPY 664
Qy 126 ----- 125
Db 665 REVOCSEFNKDIGIQGVASTNTHWVPKYANVAPNERCKLYCRLSGSAAFYLLRDKVVDG 724
Qy 126 -----GACSP----- 130
Db 725 TPCDRNGDDICVAGACMPAGCDHQLHSLRRDKCGVCGDDSSCKVYKGTFFNEQGTFGYN 784
Qy 131 ----- 130
Db 785 EVMKIPAGSANIDIRQGYNNMKEDDNYLSLAANGFLLNGHFQVSLARQOIAFQDTVL 844
Qy 131 ----- 130
Db 845 EYSGSDAIIERINGTPIRSIDYVHVLVSGSHPPDISYEMTAAPNAVIRPISSALYLW 904
Qy 131 ----TWS-----CL-----IT 137
Db 905 RVTDTWTECDRACRQOQSKLCLDMSTHROSHDRNCONVLKPKOATMCMNIDCSTWIT 964
Qy 138 ED----- 139
Db 965 EDVSSCSAKCGGQKQKRVSCVKMEGDRQTPASEHLCDRNSKPSDIASCIYDCSGRKWNY 1024
Qy 140 ----- 139
Db 1025 GEWTSCTCGSNGMHRKSYCVDDSNRRVDESICGREQKEATERECNRIPCPRWYVYGHW 1084
Qy 140 ----- 139
Db 1085 SECSRSCDGGVKMRHAQCLDAADRETHTSRCGPAQTQEHCHNEHACTWWQFGVMSDCSAK 1144
Qy 140 ----- 139
Db 1145 GDGVQYRDANCTDRHRSVLPHEHCLKMEKIITKPCRHRESCPKYKLGWSQCSVSCEDGWS 1204
Qy 140 ----- 139
Db 1205 SRRVSCVSGNGTEVDMSLCGTASDRPASHQTCNLGTCPFWRNWDWSACSVSCGIGHRETT 1264
Qy 140 ----- 139
Db 1265 TECIYREQSVDAFCGDTKMPETSTQCHLLPCTSMKPSHSPCSVTCGSGIQTRSVSCTR 1324
Qy 140 -----TC----- 141
Db 1325 GSEGTIYDEYFCDNRNTRPKKTCEKOTCDGPRVLQKLOADVPPPIRWATGPWTACSATCG 1384

```

```

Qy 142 ----- 141
Db 1385 NGTORLLKCRDHRDLDPDEYCNHLDKEVSTRNCRLRDCSYWMAEWEECPATCGTHVQQ 1444
Qy 142 ----- 141
Db 1445 SRNVTCVSAEDGGRTILKQVDCDVQKRPPTSARNCRLEPCPKGEEHIGSMIIGDWSKCSAS 1504
Qy 142 -----PD----- 143
Db 1505 CGGGWRRRSVCTSSSCDETRKPKMFDKCNELCPPLTNNSMOISPWTHCSVSCGGGVOR 1564
Qy 144 ----- 143
Db 1565 RKIWCEVLSGRKQDDIECSEIKPREORDCEMPPCRSYHYNKTTSSASMTSLSSSNNTTS 1624
Qy 144 -----LGVTI 148
Db 1625 SASASSLPILPPVVSQTSAMSACSACGCGTKRVRVVECVNPSLNTV 1672

```

Search completed: March 20, 2003, 12:40:57
Job time : 97 secs

GenCore version 5.1.4.ps 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:33:39 ; Search time 22.0541 Seconds
(without alignments)
181.441 Million cell updates/sec

Title: SEQID_15

Perfect score: 716

Sequence: 1 AVGPDPVQAHQEDTERVLT.....SPTWSCLITDGTGDLGVTI 136

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp:*

2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp:*

3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp:*

4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp:*

5: /cgn2_6/ptodata/2/iaa/PTCUS COMB.pcp:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	212.5	29.7	997	4	US-09-369-364A-7
2	184	25.7	1882	4	US-09-369-364A-13
3	178.5	24.9	967	4	US-09-130-491-2
4	176	24.6	859	4	US-09-369-364A-5
5	173.5	24.1	837	4	US-09-122-126B-2
6	170	23.7	245	4	US-09-369-364A-11
7	159	22.2	1205	4	US-09-491-522-11
8	157	21.9	566	4	US-09-491-522-7
9	157	21.9	1211	4	US-09-491-522-5
10	154.5	21.6	874	4	US-09-369-364A-15
11	149.5	20.9	930	4	US-09-369-364A-2
12	147.5	20.6	491	4	US-09-930-872-2
13	147.5	20.6	1224	4	US-09-930-872-4
14	147	20.5	905	4	US-09-369-364A-9
15	141.5	19.8	930	4	US-09-122-126B-15
16	133.5	18.6	1081	4	US-09-369-364A-17
17	133	18.6	608	4	US-09-130-491-13
18	72	10.1	531	2	US-08-975-114A-4
19	72	10.1	531	4	US-08-849-281A-4
20	68.5	9.6	536	2	US-08-975-114A-5
21	68.5	9.6	536	4	US-08-849-281A-3
22	66.5	9.3	273	2	US-07-857-224B-73
23	66.5	9.3	835	4	US-09-291-819-2
24	66.5	9.3	2539	4	US-09-413-814-42
25	65.5	9.1	488	4	US-08-939-309-10
26	65.5	9.1	568	4	US-08-939-309-4
27	65	9.1	315	4	US-09-134-001C-4141

28 65 9.1 793 1 US-08-015-985-3 Sequence 3, Appli
29 64.5 9.0 568 4 US-09-238-173-2 Sequence 2, Appli
30 63 8.8 739 3 US-08-510-648B-33 Sequence 33, Appl
31 62.5 8.7 797 1 US-08-453-695A-112 Sequence 112, App
32 62.5 8.7 797 1 US-08-268-161A-112 Sequence 112, App
33 62.5 8.7 797 2 US-08-453-702A-112 Sequence 112, App
34 62.5 8.7 797 2 US-09-099-639-112 Sequence 112, App
35 62.5 8.7 797 5 PCT-US95-08071-112 Sequence 112, App
36 62.5 8.7 3672 2 US-08-822-445-12 Sequence 12, Appl
37 62.5 8.7 3672 4 US-09-396-540-12 Sequence 12, Appl
38 62.5 8.7 3801 2 US-08-822-445-10 Sequence 10, Appl
39 62.5 8.7 3801 4 US-09-396-540-10 Sequence 10, Appl
40 61.5 8.6 568 4 US-08-939-309-2 Sequence 2, Appli
41 61.5 8.6 739 1 US-07-618-946B-22 Sequence 22, Appli
42 61.5 8.6 814 1 US-07-618-946B-23 Sequence 23, Appl
43 61 8.5 585 1 US-08-477-674-10 Sequence 10, Appl
44 61 8.5 585 1 US-08-473-791-10 Sequence 10, Appl
45 61 8.5 585 2 US-08-316-714-10 Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-09-369-364A-7

; Sequence 7, Application US/09369364A

; Patent No. 6391610

; GENERAL INFORMATION:

; APPLICANT: Apte, Suneel

; APPLICANT: Hurskainen, Tiina L.

; APPLICANT: Hirohata, Satoshi

; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases

; FILE REFERENCE: 26473/4007/10-30-00

; CURRENT APPLICATION NUMBER: US/09/369,364A

; CURRENT FILING DATE: 1999-08-06

; NUMBER OF SEQ ID NOS: 31

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 7

; LENGTH: 997

; TYPE: PRT

; ORGANISM: Homo sapiens ADAMTS-7

; US-09-369-364A-7

Query Match 29.7%; Score 212.5; DB 4; Length 997;

Best Local Similarity 37.5%; Pred. No. 6.5e-18;

Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;

Oy 11 QEDTERVYLVNIGAEELLRDPGLGAQFRVHLVXNVILTEPEGAPNITANLTSSLVCG 70

Db 260 QPQVESYVLTIMNVAGLPHDPSIGNPIHITIVRLVLEDEEDLKITHAQLTKSFCK 319

Oy 71 WSOITNPEDDTDCHADLVLYITRFDLELPGDNR--QVRGVTLGGACSPWMSCLITD 128

Db 320 WQKSINNKGDAHLPHDHTAILLTRKDL-CAAMNRPCTGLSHVAGMCOPIHRSINEDT 378

Oy 129 GFDLGVTI 136

Db 379 GLPLAFTV 386

RESULT 2

US-09-369-364A-13

; Sequence 13, Application US/09369364A

; Patent No. 6391610

; GENERAL INFORMATION:

; APPLICANT: Apte, Suneel

; APPLICANT: Hurskainen, Tiina L.

; APPLICANT: Hirohata, Satoshi

; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases

; FILE REFERENCE: 26473/4007/10-30-00

; CURRENT APPLICATION NUMBER: US/09/369,364A

; CURRENT FILING DATE: 1999-08-06

; NUMBER OF SEQ ID NOS: 31

Query Match 21.9%; Score 157; DB 4; Length 566;
Best Local Similarity 31.3%; Pred. No. 3.1e-11;
Matches 41; Conservative 24; Mismatches 60; Indels 6; Gaps 3;

QY 2 VGPVFAQH-QEDTERVLTNLNIGAEALLRDPGLGAFRVLHVLMVILTEPEGAPNI-TA 59
DB 274 VDDSVQFHGKEHVQKYLTLNINVEIYHDESLGAINVVLVRIILLSYGKMSLIEIG 333

QY 60 NLTSSLSVCGWSTINPEDDTPGHADLVLYITRFDLEPDGNGRQVRGVTVQLGGACSP 119
DB 334 NPSQLENVCRWAYLOOKPDTGHDYHDAIFLTRQDF----GPSGMOGVAPVTGMCHPV 389

QY 120 WSLTITDTCF 130
DB 390 RSCTLNHEDGF 400

RESULT 9
US-09-491-522-5
; Sequence 5, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Colige, Alain
; APPLICANT: Lapierre, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B.
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1211 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-491-522-5

Query Match 21.9%; Score 157; DB 4; Length 1211;
Best Local Similarity 31.3%; Pred. No. 9.5e-11;
Matches 41; Conservative 24; Mismatches 60; Indels 6; Gaps 3;

QY 2 VGPVFAQH-QEDTERVLTNLNIGAEALLRDPGLGAFRVLHVLMVILTEPEGAPNI-TA 59
DB 274 VDDSVQFHGKEHVQKYLTLNINVEIYHDESLGAINVVLVRIILLSYGKMSLIEIG 333

QY 60 NLTSSLSVCGWSTINPEDDTPGHADLVLYITRFDLEPDGNGRQVRGVTVQLGGACSP 119

DB 334 NPSQLENVCRWAYLOOKPDTGHDYHDAIFLTRQDF----GPSGMOGVAPVTGMCHPV 389
QY 120 WSLTITDTCF 130
DB 390 RSCTLNHEDGF 400

RESULT 10
US-09-369-364A-15
; Sequence 15, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 15
; LENGTH: 874
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-9
US-09-369-364A-15

Query Match 21.6%; Score 154.5; DB 4; Length 874;
Best Local Similarity 27.5%; Pred. No. 1.2e-10;
Matches 36; Conservative 26; Mismatches 58; Indels 11; Gaps 2;

QY 10 HQEDTERVLTNLNIGAEALLRDPGLGAFRVLHVLMVILTEPEGAPNITANLTSSLSVC 69
DB 148 HGANLQHYLTLMISIVASYIKOSSICNLINIVNLVHNHNEOGPYINFNAOTLLKNFC 207
QY 70 GWSQTIN---PEDDTPGHADLVLYITRFDLEPDGNGRQVRGVTVQLGGACSP 125
DB 208 QWQHSKNYLGIGIHDFT-----AVLTREDICRAQDKCDTLGLAELGTICDPYRSCSIS 260
QY 126 EDTGFDLGVTI 136
DB 261 EDSGLSTAF 271

RESULT 11
US-09-369-364A-2
; Sequence 2, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 930
; TYPE: PRT
; ORGANISM: mus musculus ADAMTS-5
US-09-369-364A-2

Query Match 20.9%; Score 149.5; DB 4; Length 930;
Best Local Similarity 28.7%; Pred. No. 5.8e-10;
Matches 35; Conservative 22; Mismatches 64; Indels 1; Gaps 1;

QY 15 ERYVLNINIGAEALLRDPGLGAFRVLHVLMVILTEPEGAPNITANLTSSLSVC 74
DB 288 QHYLLTMASIANRLYSHASIHENIRLAVVVKVVVLTOKTSLEVSKNAAATLLKNFCKWQHQ 347

Db 288 QHYLLTLASIANRLYSHASIENHIRLAVKVVVLGDKDKSLEVSKNAATTLKNFCKWQH 347
Qy 75 INPEDDTPGHADLVLYITREDLELPDGNRQVRGVTOLGACSPWCLITETGFDLGV 134
Db 348 HNQLCDDHEEHYDAAILFTREDL-CGHHSCDTLGMADVGTICSPERSCAVIEDDGLHAAP 406
Qy 135 TI 136
Db 407 TV 408

Search completed: March 20, 2003, 12:36:05
Job time : 26.0541 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:32:34 ; Search time 24.8108 Seconds
(without alignments)
526.959 Million cell updates/sec

Title: SEQID_15
Perfect score: 716
Sequence: 1 AVGPVFAHQEDTERVYL.....SPTWSCLITDGFGLGVTI 136
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	201	28.1	2165	2 T21371	hypothetical prote
2	177.5	24.8	951	2 T00017	gene ADAMTS-1 prot
3	172.5	24.1	837	2 T00355	hypothetical prote
4	159	22.2	1205	2 T18517	procollagen N-endo
5	112.5	15.7	860	2 T16892	hypothetical prote
6	85	11.9	957	2 T15976	hypothetical prote
7	75.5	10.5	228	2 C84824	hypothetical prote
8	75.5	10.5	508	2 T22836	hypothetical prote
9	73.5	10.3	1444	2 T18856	angiogenesis inhib
10	72	10.1	531	2 JN0586	beta-1,4-mannosyl-
11	71.5	10.0	723	2 T14765	hypothetical prote
12	71	9.9	236	2 G82052	ribulose-phosphate
13	71	9.9	443	2 H82957	probable glutamine
14	71	9.9	726	2 T08769	hypothetical prote
15	70.5	9.8	266	2 E71230	hypothetical prote
16	70.5	9.8	388	2 T36402	probable monooxyge
17	70.5	9.8	475	2 F70972	probable amidase -
18	70.5	9.8	539	2 T42624	probable translati
19	70.5	9.8	1403	2 T11583	probable translati
20	70	9.8	369	2 D81196	conserved hypothet
21	70	9.8	375	2 A81086	conserved hypothet
22	70	9.8	375	2 G81857	hypothetical prote
23	69.5	9.7	161	2 E83541	peptidyl-prolyl ci
24	69.5	9.7	210	2 T22344	hypothetical prote
25	69.5	9.7	305	2 D69199	conserved hypothet
26	69	9.6	157	2 AH3460	arsenate reductase
27	69	9.6	352	2 G82719	alcohol dehydrogen
28	69	9.6	875	2 I59350	karyopherin beta -
29	69	9.6	876	2 S66288	nuclear pore-targe

30	68.5	9.6	536	2 A43415	beta-1,4-mannosyl-
31	68.5	9.6	536	2 JC4362	beta-1,4-mannosyl-
32	68.5	9.6	706	2 A86350	F8K7.10 protein -
33	68.5	9.6	861	2 S51350	KAP95 protein - ye
34	68	9.5	225	2 AH0019	ribulose-phosphate
35	68	9.5	372	2 A41792	homeotic protein e
36	68	9.5	430	2 AE0168	serine-tRNA ligase
37	68	9.5	876	2 I52907	importin beta chai
38	68	9.5	1706	2 B75633	probable RNA helic
39	68	9.5	4845	2 T31067	BIR repeat contain
40	67.5	9.4	251	2 G72118	conserved hypothet
41	67.5	9.4	251	2 H86504	CT021 hypothetical
42	67.5	9.4	536	2 AF0523	probable multicopp
43	67.5	9.4	647	2 G70733	probable htpG prot
44	67.5	9.4	655	2 F83597	hypothetical prote
45	67.5	9.4	674	2 AC1913	two-component sens

ALIGNMENTS

RESULT 1
T21371
hypothetical protein F25H8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21371; T24896
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19413
A:Accession: T21371
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WIL>
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone F25H8
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19949
A:Accession: T24896
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WT2>
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone T13H10
A:Gene: CESP:F25H8.3
A:Map position: 4
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/2

Query Match 28.1%; Score 201; DB 2; Length 2165;
Best Local Similarity 33.6%; Pred. No. 1.4e-12;
Matches 44; Conservative 17; Mismatches 70; Indels 0; Gaps 0;
QY 6 VFAHQEDTERVYLNINAGELLRPSLGAFRHLVKWVILTEPEGAPNITANLTSSL 65
DB 292 MYEYHGRSLDYVLTFSFVASIYHQSLRASINVVVKLVILKKTENAGPRITONAQOTL 351
QY 66 LSCVGNSTOINPDDTDTPGHADLVLYITFDLELPDGNRQVRGVGTOLGGACSPWTSCLIT 125
DB 352 QDFCRWQYINPDSSVQHVHVAILLTRKDCRSQKCDTGLAELGTWCDMQKSCAII 411
QY 126 EDTGFDLGVTI 136
DB 412 EDNGLSAFTI 422

RESULT 2
T00017
gene ADAMTS-1 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00017

RESULT 8

Db 355 CSLVE--GLHLG 364

Search completed: March 20, 2003, 12:35:34
Job time : 28.8108 secs

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:32:11 ; Search time 16.5405 Seconds
(without alignments)
341.028 Million cell updates/sec

Title: SEQID_15

Perfect score: 716

Sequence: 1 AVGPDVFOAQHEDTERVYLT.....SPTWSCLITDGTGDLGVTI 136

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	212.5	29.7	997	1 AT57_HUMAN	Q9ukp4 homo sapien
2	211.5	29.5	1593	1 AT12_HUMAN	P58397 homo sapien
3	187.5	26.2	1629	1 AT59_HUMAN	Q9p2n4 homo sapien
4	178.5	24.9	967	1 AT51_HUMAN	Q9uh18 homo sapien
5	177.5	24.8	967	1 AT51_RAT	Q9uhq1 rattus norv
6	177.5	24.8	968	1 AT51_MOUSE	P97857 mus musculu
7	176	24.6	860	1 AT56_HUMAN	Q9ukp5 homo sapien
8	172.5	24.1	837	1 AT54_HUMAN	Q75173 homo sapien
9	172	24.0	1077	1 AT10_HUMAN	Q9h324 homo sapien
10	170	23.7	890	1 AT58_HUMAN	Q9up79 homo sapien
11	168.5	23.5	630	1 AT54_RAT	Q9esp7 rattus norv
12	159	22.2	1205	1 AT52_BOVIN	P79331 b adamts-2
13	157	21.9	1211	1 AT52_HUMAN	Q95450 h adamts-2
14	153	21.4	1205	1 AT53_MOUSE	O15072 homo sapien
15	148.5	20.7	930	1 AT55_MOUSE	Q9r001 mus musculu
16	147	20.5	905	1 AT58_MOUSE	P57110 mus musculu
17	141.5	19.8	930	1 AT55_HUMAN	Q9una0 homo sapien
18	82	11.5	1862	1 GSAB_PICPA	Q9hfr4 pichia past
19	72	10.1	531	1 GNT3_HUMAN	Q09327 h beta-1,4-
20	70.5	9.8	475	1 AM14_MYCTU	O50404 mycobacteri
21	70.5	9.8	1403	1 YDF3_SCHPO	Q10475 schizosacch
22	70	9.8	359	1 PTN7_RAT	P49445 rattus norv
23	69	9.6	875	1 IMB1_RAT	P52296 rattus norv
24	69	9.6	876	1 IMB1_MOUSE	P70168 mus musculu
25	68.5	9.6	536	1 GNT3_MOUSE	Q10470 m beta-1,4-
26	68.5	9.6	536	1 GNT3_RAT	Q02527 r beta-1,4-
27	68.5	9.6	861	1 IMB1_YEAST	Q06142 saccharomyc
28	68	9.5	372	1 HMEN_BOWMO	P27609 bombyx mori
29	68	9.5	876	1 IMB1_HUMAN	Q14974 homo sapien
30	67.5	9.4	536	1 CUEO_SALTI	Q82961 salmonella
31	67.5	9.4	536	1 CUEO_SALTY	Q82962 salmonella
32	67.5	9.4	647	1 HTPG_MYCTU	O50667 mycobacteri
33	67.5	9.4	801	1 CDB6_HUMAN	Q9un66 homo sapien

34	67	9.4	534	1	EPAL_MOUSE
35	67	9.4	575	1	VG05_BPT4
36	67	9.4	789	1	AD07_RAT
37	67	9.4	859	1	YOD3_CABEL
38	66.5	9.3	184	1	YGB8_YEAST
39	66.5	9.3	467	1	DNPE_SCHPO
40	66	9.2	504	1	CPD9_MOUSE
41	66	9.2	1083	1	MAN1_YEAST
42	65.5	9.1	451	1	VDR_CHICK
43	65.5	9.1	616	1	RRPO_PLRV1
44	65.5	9.1	616	1	RRPO_PLRVW
45	65.5	9.1	1077	1	YBK1_YEAST

ALIGNMENTS

RESULT 1
AT57_HUMAN
ID AT57_HUMAN STANDARD; PRT; 997 AA.
AC Q9UKP4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloprotease)
DE with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7).
GN ADAMTS7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9395124; PubMed=10464288;
RA Hurekainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, Novel Members of a New Family of Zinc Metalloproteases";
RL J. Biol. Chem. 274:25555-25563(1999).
CC -|- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).
CC -|- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS.
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -|- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; AF140675; RAD56358.1; -;
CC HSP; P15167; IATL.
CC MEROPS; M12.231; -;
CC Genew; HGNC:223; ADAMTS7.
CC MIM; 605009; -;
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR00130; Zn_M12Bpropep.
CC Pfam; PF00090; tsp_1; -;
CC Pfam; PF01421; Reprolysin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC SMART; SM00209; TSP1; 2.

```
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS0215; ADAM_MEPRO; 1.
DR PROSITE: PS0092; TSP1_1.
DR PROSITE: PS0427; DISINTEGRIN_1; FALSE NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
FT Repeat; Extracellular matrix.
FT SIGNAL 1 27
FT PROPEP 28 232 BY SIMILARITY.
FT CHAIN 233 997 ADAMTS-7.
FT SITE 204 204 CYSTEINE SWITCH (POTENTIAL).
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 389 389 BY SIMILARITY.
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 462 537 DISINTEGRIN-LIKE.
FT DOMAIN 538 594 TSP TYPE-1 1.
FT DOMAIN 595 697 CYS-RICH.
FT DOMAIN 698 914 SPACER.
FT DOMAIN 915 990 TSP TYPE-1 2.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED03FC104 CRC64;

Query Match 29.7%; Score 212.5; DB 1; Length 997;
Best Local Similarity 37.5%; Pred No. 7e-14;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;

Qy 11 QEDTERVLTNINICAEELRLDPSLQAFRVHLVKVILTEPGANITANLTSSLLSVCVG 70
Db 260 QPQVESYVLTMMVMVAGLPHDPSIGNPIHITIVRLVLLLEDEEDLKITHADNTUKSPCK 319

Qy 71 WSQTNPEDDTDPGHADLVLYITRDELPLDGNR--QVRGVTLGAGCSPTWSCLITEDT 128
Db 320 WOKSINMGDAHPLHDDTAILTRKDL-CAAMNRPCETGLSHVAGMCPHRSCSINEDT 378

Qy 129 GFDLGVTI 136
Db 379 GLPLAFTV 386

RESULT 2
AT12 HUMAN STANDARD; PRT; 1593 AA.
AC PS0397;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-
DE TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=21264577; PubMed=11279086;
RA Cal S., Arqueles J.M., Fernandez P.L., Lopez-Otin C.;
RT Identification, characterization, and intracellular processing of
RT ADAM-TS12, a novel human disintegrin with a complex structural
RT organization involving multiple thrombospondin-1 repeats.;
RL J. Biol. Chem. 276:17932-17940(2001).
CC -!- COPACITOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
```

```
CC SIMILARITY).
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -!- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
CC DOMAINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ250725; CAC20419.1; .
CC Genew; HGNC.14605; ADAMTS12.
CC MIM; 606184; .
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reptolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; Zn_MTpeptdae.
CC Pfam; PF00090; tsp_1; 6.
CC Pfam; PF01421; Reptolysin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC SMART; SM00209; TSP1; 8.
CC PROSITE; PS0215; ADAM_MEPRO; 1.
CC PROSITE; PS00437; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE; PS00092; TSP1; 2.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
Repeat; Extracellular matrix.
FT SIGNAL 1 25
FT PROPEP 26 240 BY SIMILARITY.
FT CHAIN 241 1593 ADAMTS-12.
FT DOMAIN 465 544 DISINTEGRIN-LIKE.
FT DOMAIN 545 596 TSP TYPE-1 1.
FT DOMAIN 597 700 CYS-RICH.
FT DOMAIN 701 826 SPACER 1.
FT DOMAIN 827 881 TSP TYPE-1 2.
FT DOMAIN 886 943 TSP TYPE-1 3.
FT DOMAIN 947 995 TSP TYPE-1 4.
FT DOMAIN 996 1315 SPACER 2.
FT DOMAIN 1316 1364 TSP TYPE-1 5.
FT DOMAIN 1367 1423 TSP TYPE-1 6.
FT DOMAIN 1426 1471 TSP TYPE-1 7.
FT DOMAIN 1426 1471 TSP TYPE-1 8.
FT DOMAIN 302 305 POLY-GLU.
FT SITE 208 208 CYSTEINE SWITCH (POTENTIAL).
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 393 393 BY SIMILARITY.
FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 685 685 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 790 790 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 951 951 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1275 1275 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1300 1300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1320 1320 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1503 1503 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1593 AA; 177545 MW; 07F9F48E63BD83A3 CRC64;
```


Oy 128 TGFOLGVTI 136
 Db 424 SGLSTAFTI 432

RESULT 4
 AT51 HUMAN
 ID AT51 HUMAN STANDARD; PRT: 967 AA.
 AC 09UH18; Q9UP80; Q9UH83; Q9P2K0; Q9NSJ8;
 DT 30-MAY-2000 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).
 GN ADAMTS1 OR METH1 OR KIAA1346.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Casas C., Pritchard M.A., Estivill X., Arbones M.L.;
 RT "Cloning, characterization and mapping on human chromosome 21 of the
 RT orthologue of murine Adamts-1.";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. AND FUNCTION.
 RC TISSUE=Heart;
 RX MEDLINE=99367466; PubMed=10438512;
 RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
 RA Lombardo M., Iruela-Arispe M.L.;
 RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
 RT family of proteins with angio-inhibitory activity.";
 RL J. Biol. Chem. 274:23349-23357(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Endothelial cells;
 RX MEDLINE=20247184; PubMed=10785405;
 RA Glenske J., Schmitt A.O., Pilarsky C., Hinzmann B., Weiss B.,
 RA Rosenthal A., Thierauch K.H.;
 RT "Differential gene expression by endothelial cells in distinct
 RT angiogenic states.";
 RL Eur. J. Biochem. 267:2820-2830(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes.XVI.
 RT The complete sequences of 150 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:65-73(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
 RA Okhi M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
 RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyma S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordliek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [6]
 RP SEQUENCE OF 418-967 FROM N.A.
 RC TISSUE=Melanoma;
 BLUM H., Bauersachs S., Mewes H.-W., Well B., Wiemann S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
 CC INVOLVED IN ITS TURNOVER (BY SIMILARITY). HAS ANGIOGENIC INHIBITOR
 CC ACTIVITY. ACTIVE METALLOPROTEASE WHICH MAY BE ASSOCIATED WITH
 CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
 CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE.
 CC CATALYTIC ACTIVITY: CLEAVES AGGECAN AT THE 1938-GLU-|-LEU-1939
 CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (by similarity).
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF170084; AAF15317.1; -;
 DR EMBL; AF060152; AAD48080.1; ALT_INIT.
 DR EMBL; AF207664; AAF23772.1; -;
 DR EMBL; AB037767; BAA92584.1; ALT_INIT.
 DR EMBL; AP001697; BAA95502.1; -;
 DR EMBL; AL162080; CAB82413.1; -;
 DR MEROPS; M12.222; -;
 DR Genew; HGNC:217; ADAMTS1.
 DR MIM; 605174; -;
 DR InterPro; IPR001762; Disintegrin.
 DR InterPro; IPR002870; Pep M12B propep.
 DR InterPro; IPR001590; Repolysin.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR000130; Zn_Mtpeptdse.
 DR Pfam; PF00090; tsp_1; 6;
 DR Pfam; PF01421; Repolysin; 2.
 DR SMART; SM00209; TSP1; 3;
 DR SMART; SM00215; ADAM_MEPRO; 1; FALSE_NEG.
 DR PROSITE; PS00427; DISINTEGRIN_1;
 DR PROSITE; PS50092; TSP1; 2;
 DR PROSITE; PS00142; ZINC_PROTEASE; 1;
 DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Extracellular matrix; Heparin-binding.
 FT SIGNAL 1 49
 FT PROPEP 50 252
 FT CHAIN 253 967
 FT SITE 198 198
 FT METAL 401 401
 FT ACT_SITE 402 402
 FT METAL 405 405
 FT METAL 411 411
 FT DOMAIN 476 559
 FT DOMAIN 560 616
 FT DOMAIN 617 724
 FT DOMAIN 725 849
 FT DOMAIN 850 908
 FT DOMAIN 909 967
 FT DOMAIN 843 846
 FT CARBOHYD 547 547
 FT CARBOHYD 720 720
 FT CARBOHYD 764 764
 FT CONFLICT 227 227
 FT CONFLICT 468 468
 FT CONFLICT 561 561
 FT SEQUENCE 967 AA; 105383 MW; C189389324741ED1 CRC64;


```

Query Match      24.9%; Score 178.5; DB 1; Length 967;
Best Local Similarity 28.9%; Pred. No. 2.1e-10;
Matches 37; Conservative 26; Mismatches 60; Indels 5; Gaps 2;

QY 10 HQSDTERVYLTNLNIGAEILLRDPGLGQAFRVHLVWVILTEPEGAPNITANLTSSLLSVC 69
DB 274 HGSGLKHLLTLFSAARLYKHPSIRNSISLVVVKILVIHDEQKGPVTSNAALTNRFC 333
QY 70 GWSOTINPEDTDPGHADLVLYITRFDELPLDGNR--QVRGVTLGGACGPTWSCLITD 127
DB 334 NWQKHNPSPDRDAEHYDTAILFTRQDL---CGSQTCDTLGMADVGVCDPSRCSVIED 390
QY 128 TGFDLGVT 135
DB 391 DGLQAFT 398

RESULT 5
ATSL_RAT
ID ATSL_RAT STANDARD; PRT; 967 AA.
AC O9MU01; O9ER11;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN ADAMTS1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
ON NCBI_TaxID=10116;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RA Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,
RA Little S.P.;
RT "Induction of a disintegrin and metalloproteinase with the
RT thrombospondin type 1 motif (ADAMTS)",
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 18-967 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=20304099; PubMed=10847486;
RA Diamantis I., Luehli M., Hoesli M., Reichen J.;
RA "Cloning of the rat ADAMTS-1 gene and its down regulation in
RA endothelial cells in cirrhotic rats.";
RL Liver 20:165-172(2000).
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER, HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU-1-LEU-1684
CC SITE WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COPACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -1- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM
CC CIRRHOTIC LIVER.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isab-sib.ch/announce/
or send an email to license@isab-sib.ch).
-----
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isab-sib.ch/announce/
CC CC or send an email to license@isab-sib.ch).
CC CC -----
DR EMBL; AF149118; AAD34012.1; -;
DR MEROPS; M12.222; -;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTPeptide.
DR Pfam; PF00090; tep_1; 6.
DR Pfam; PF01421; Reprolysin; 2.
DR Pfam; PF01562; Pep M12B_propep; 2.
DR SMART; SM00209; TSP1_3;
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00442; ZINC_PROTEASE; 1.
DR PROSITE; PS00092; TSP1; 2.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 54
FT PROPEP 55 252
FT CHAIN 253 967
FT SITE 205 205
FT METAL 401 401
FT ACT_SITE 402 402
FT METAL 405 405
FT METAL 411 411
FT DOMAIN 476 558
FT DOMAIN 559 615
FT DOMAIN 616 724
FT DOMAIN 725 857
FT DOMAIN 858 907
FT DOMAIN 908 967
FT DOMAIN 134 198
FT CARBOHYD 547 547
FT CARBOHYD 720 720
FT CARBOHYD 764 764
FT CARBOHYD 782 782
FT CARBOHYD 945 945
FT CONFLICT 21 21
FT CONFLICT 26 31
FT CONFLICT 49 49
FT CONFLICT 72 72
FT CONFLICT 79 79
FT CONFLICT 249 249
FT CONFLICT 262 265
FT CONFLICT 607 607
FT CONFLICT 936 936
FT CONFLICT 962 962
SQ SEQUENCE 967 AA; 105705 MW; F93C864F6DCDB4CF CRC64;

Query Match      24.8%; Score 177.5; DB 1; Length 967;
Best Local Similarity 28.6%; Pred. No. 2.6e-10;
Matches 36; Conservative 23; Mismatches 66; Indels 1; Gaps 1;

QY 10 HQSDTERVYLTNLNIGAEILLRDPGLGQAFRVHLVWVILTEPEGAPNITANLTSSLLSVC 69
DB 274 HGSGLKHLLTLFSAARLYKHPSIRNSISLVVVKILVIHDEQKGPVTSNAALTNRFC 333
QY 70 GWSOTINPEDTDPGHADLVLYITRFDELPLDGNRQVRGVTLGGACGPTWSCLITDTC 129
DB 334 NWQKHNPSPDRDAEHYDTAILFTRQDL---CGSHTCDTLGMADVGVCDPSRCSVIEDDG 392
QY 130 FDLGVT 135
DB 393 LQAFT 398

RESULT 6
ATSL_MOUSE

```

ID	ATSL MOUSE	STANDARD;	PRT;	968 AA.	CC	FRAMESHIFT IN POSITION 7.
AC	P97857; O54768;				CC	-----
DT	30-MAY-2000 (Rel. 39, Created)				CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
DT	16-OCT-2001 (Rel. 40, Last sequence update)				CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
DT	15-JUN-2002 (Rel. 41, Last annotation update)				CC	the European Bioinformatics Institute. There are no restrictions on its
DE	ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase				CC	use by non-profit institutions as long as its content is in no way
DE	with thrombospondin motifs 1) (ADAM-TS1) (ADAM-TS1)				CC	modified and this statement is not removed. Usage by and for commercial
GN	ADAMTS1				CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
OS	Mus musculus (Mouse)				CC	or send an email to license@isb-sib.ch)
OC	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				CC	-----
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				CC	ENBL; A8001735; BAA24501.1; ALT INIT.
OX	NCBI_TaxID=10090;				DR	ENBL; D67076; BAA11088.1; ALT_FRAME.
RN	[1]				DR	MEROPS; M12.222; -
RP	SEQUENCE FROM N.A.				DR	MGD; MGI:109249; Adamts1.
RC	STRAIN=129/SVJ;				DR	InterPro; IPR001762; Disintegrin.
RX	MEDLINE=99110583; PubMed=9441751;				DR	InterPro; IPR002870; Pep.M12B_propep.
RA	Kuno K., Lizasa H., Ohno S., Matsushima K.;				DR	InterPro; IPR001590; Repolysin.
RA	"The exon/intron organization and chromosomal mapping of the mouse				DR	InterPro; IPR000884; TSPI.
RT	ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";				DR	InterPro; IPR000130; Zn.MTpeptidase.
RL	Genomics 46:466-471(1997).				DR	Pfam; PF00090; tsp.1; 3.
RN	[2]				DR	Pfam; PF01421; Repolysin; 1.
RP	SEQUENCE FROM N.A.				DR	Pfam; PF01562; Pep.M12B_propep; 1.
RP	MEDLINE=97150761; PubMed=9995297;				DR	SMART; SM00209; TSPI; 3.
RA	Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.;				DR	PROSITE; PS00215; ADAM_MBP; 1. FALSE_NEG.
RA	Matsushima K.;				DR	PROSITE; PS00427; DISINTEGRIN_1; 3.
RT	"Molecular cloning of a gene encoding a new type of metalloproteinase-				DR	PROSITE; PS00092; TSPI; 3.
RT	disintegrin family protein with thrombospondin motifs as an				DR	PROSITE; PS00142; ZINC_PROTEASE; 1.
RT	inflammation associated gene.";				DR	Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
RL	J. Biol. Chem. 272:556-562(1997).				KW	Repeat; Extracellular matrix; Heparin-binding.
RN	[3]				FT	SIGNAL 1 48
RP	CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.				FT	PROPEP 49 253
RP	MEDLINE=99303657; PubMed=10373500;				FT	CHAIN 254 968
RX	Kuno K., Tereshima Y., Matsushima K.;				FT	SITE 206 206
RA	"ADAMTS-1 is an active metalloproteinase associated with the				FT	METAL 402 402
RT	extracellular matrix.";				FT	ACT_SITE 403 403
RL	J. Biol. Chem. 274:18821-18826(1999).				FT	METAL 406 406
RN	[4]				FT	METAL 412 412
RP	FUNCTION.				FT	DOMAIN 477 559
RX	MEDLINE=20389568; PubMed=10930576;				FT	DOMAIN 560 617
RA	Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.;				FT	DOMAIN 618 725
RA	Ohno H., Matsushima K.;				FT	DOMAIN 726 850
RA	"ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";				FT	DOMAIN 851 909
RL	FEBS Lett. 478:241-245(2000).				FT	DOMAIN 910 968
RN	[5]				FT	DOMAIN 195 199
RP	FUNCTION, AND INDUCTION.				FT	DOMAIN 548 548
RX	MEDLINE=20243757; PubMed=10781075;				FT	CARBOHYD 721 721
RA	Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W.;				FT	CARBOHYD 765 765
RA	Richards J.S.;				FT	CARBOHYD 783 783
RT	"Progestosterone-regulated genes in the ovulation process: ADAMTS-1 and				FT	CARBOHYD 946 946
RT	cathespain L proteases.";				FT	MUTAGEN 403 403
RL	Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).				FT	CONFLICT 335 335
CC	-1- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE				FT	CONFLICT 425 425
CC	INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY				SO	SEQUENCE 968 AA; 105841 MW; 42EBDA55499FB6C1 CRC64;
CC	SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH					Query Match 24.8%; Score 177.5; DB 1; Length 968;
CC	VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER					Best Local Similarity 28.6%; Pred. No. 2.6e-10;
CC	CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY					Matches 36; Conservative 23; Mismatches 66; Indels 1; Gaps 1;
CC	SIMILARITY).					
CC	-1- CATALYTIC ACTIVITY: CLEAVES AGGECAN AT THE 1691-GLU- -LEU-1692					
CC	SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.					
CC	-1- COPACITOR: BINDS 1 ZINC ION (BY SIMILARITY).					
CC	-1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR					
CC	MATRIX.					
CC	-1- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY					
CC	INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY					
CC	LIPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA					
CC	CELLS OF PREOVULATORY FOLLICLES.					
CC	-1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT					
CC	FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.					
CC	-1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.					
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.					
CC	-1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.					
CC	-1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.					
CC	-1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A					

RESULT 7
ATSG_HUMAN

AD AC Q9UKP5; 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-6 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
GN ADAMTS6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99395124; PubMed=10464288;
RA Hurekainen T.L., Hirschata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
Zinc Metalloproteases";
RL J. Biol. Chem. 274:25555-25563 (1999).
CC -|- CAPSATOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (by similarity).
CC -|- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN PLACENTA AND BARELY
DETECTABLE IN A NUMBER OF OTHER TISSUES.
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -|- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
SIMILARITY).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation.
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF140674; AAD56357.1; -.
DR HSP; P34179; 11AG.
DR MEROPS; M12.230; -.
DR Genew; HGNC:222; ADAMTS6.
DR MIM; 605008;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF00090; tsp1; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS0215; ADAM_MEPPO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00992; TSP1; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 244 BY SIMILARITY.
FT CHAIN 245 860 ADAMTS-6.
FT METAL 403 403 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 404 404 BY SIMILARITY.
FT METAL 407 407 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 413 413 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 453 509 DISINTEGRIN-LIKE.
FT DOMAIN 510 566 TSP TYPE-1 1.
FT DOMAIN 567 668 CYS-RICH.
FT DOMAIN 669 795 SPACER.
FT DOMAIN 796 852 TSP TYPE-1 2.
FT DOMAIN 68 71 POLY-ARG.

FT DOMAIN 662 665 POLY-GLY.
FT CARBOHYD 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 676 676 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 843 843 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 860 AA; 97098 MW; E57213015DECB2C5 CRC64;
Query Match 24.6%; Score 176; DB 1; Length 860;
Best Local Similarity 33.8%; Pred. No. 3.3e-10;
Matches 47; Conservative 23; Mismatches 51; Indels 18; Gaps 6;
Oy 11 QEDTERVYVLTNLTGAEALLRDPISLGAOPRVHLVKVILTEPEGAPNITAN--LTSSLSLV 68
Db 268 RKDIEHYLSVMNIVAKLYRDSLSGVNVIIVARLIVLTQDQ--PNLEINHAKSLDSF 325
Oy 69 CGMSQTI---NPEDDTDP---GHADLVLYITREDL---ELPDGNRQVRGVTOLGAGCS 117
Db 326 CKWOKSLSHOSDGNTPENGIAHHDNAVLITRYDICTYKNKPCG---TLGLASVAGMCE 382
Oy 118 PTWCLITEDTGFDLGVTI 136
Db 393 PERSCSINEDIGLSAFTI 401
RESULT 8
ATSA HUMAN STANDARD; PRT; 837 AA.
AC OVS173; G9UN83;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
DE (ADMP-1).
GN ADAMTS4 OR KIAA0688.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9840380; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9286303; PubMed=10356395;
RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
RA Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
RA Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,
RA Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
RA Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
RA Trzaskos J.M., Arner E.C.;
RT "Purification and cloning of aggrecanase-1: a member of the ADAMTS
family of proteins";
RL Science 284:1664-1666(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Savaji Y., Nagase H., Saklatvala J., Clark A.R.;
RT "ADAMTS-4 genomic locus";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=20400518; PubMed=10827174;
RA Tortorella M.D., Pratta M.A., Liu R.-Q., Abbaszade I., Ross H.,
RA Burn T.C., Arner E.C.;

RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
 RT aggrecan substrate recognition and cleavage.",
 RL J. Biol. Chem. 275:25791-25797(2000).
 CC -1- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
 CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
 CC DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE
 CC CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN
 CC ALZHEIMER'S DISEASE.
 CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393
 CC site.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED
 CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
 CC -1- INDUCTION: BY INTERLEUKIN-1.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
 CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO AS ADAMTS2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AB014588; BAA31663.1; -;
 CC EMBL: AF148213; AAD41494.1; -;
 CC EMBL: AY044847; AAL02262.1; -;
 CC MEROPS: M12.221; -;
 CC Genew: HGNC:220; ADAMTS4.
 CC MIM: 603876; -;
 CC InterPro: IPR001762; Disintegrin.
 CC InterPro: IPR001590; Reprolysin.
 CC InterPro: IPR000884; TSPI.
 CC InterPro: IPR000130; Zn_M12peptidase.
 CC Pfam: PF00090; tsp_1; 1;
 CC Pfam: PF01421; Reprolysin; 1.
 CC SMART: SM00209; TSPI; 1.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 CC PROSITE: PS0215; ADAM_MEPRO; 1.
 CC PROSITE: PS00092; TSPI; 1.
 CC PROSITE: PS00427; DISINTEGRIN_1; FALSE NEG.
 CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 CC Extracellular matrix.
 CC SIGNAL 1 51
 CC PROPEP 52 212 POTENTIAL.
 CC CHAIN 213 837 ADAMTS-4.
 CC SITE 194 194 CYS-STEINE SWITCH (POTENTIAL).
 CC METAL 361 361 ZINC (CATALYTIC) (BY SIMILARITY).
 CC ACT_SITE 362 362 BY SIMILARITY.
 CC METAL 365 365 ZINC (CATALYTIC) (BY SIMILARITY).
 CC METAL 371 371 ZINC (CATALYTIC) (BY SIMILARITY).
 CC DOMAIN 437 519 DISINTEGRIN-LIKE.
 CC DOMAIN 520 576 TSP TYPE-1.
 CC DOMAIN 577 685 CYS-RICH.
 CC DOMAIN 686 837 SPACER.
 CC DOMAIN 247 252 POLY-ALA.
 CC CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CONFLICT 77 77 A -> T (IN REF. 1).
 CC CONFLICT 626 626 R -> Q (IN REF. 3).
 CC CONFLICT 682 682 G -> R (IN REF. 3).
 CC SEQUENCE 837 AA; 90224 MW; 5DF9C9AC13DF41F CRC64;
 Query Match 24.1%; Score 172.5; DB 1; Length 837;
 Best Local Similarity 30.2%; Pred. No. 7.2e-10;
 Matches 38; Conservative 20; Mismatches 67; Indels 1; Gaps 1;

QY 10 HQSDTERVYLTNLNIGAEILLRDPISLGAQFRVHLVKWVILTEPEGAPNITANLTSSLLSVC 69
 Db 234 HGAGLKRYLTVTNVAARAKAFKPSINPVLSVTRVLVILSGEGEQVGPSSAAQTLSFC 293
 QY 70 GMSQTINPEDDTPDGHADVLVITRFDLPELPGNRQVRGVTQLGACSPWMSCLITDTC 129
 Db 294 AMQGLNTPEDSDPDHFDTAILFTRODL-CGVSTCDTLGADVTVCDPARSCAIVEDDG 352
 QY 130 FDLGVT 135
 Db 353 LOSAFT 358
 RESULT 9
 AT10_HUMAN STANDARD; PRT; 1077 AA.
 ID AT10_HUMAN
 AC Q9H324;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs 10) (ADAM-TS10) (Fragment).
 GN ADAMTS10.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Apte S.S.;
 RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
 RT thrombospondin type I repeats";
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (By
 CC similarity).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF163762; AAG35563.1; -;
 CC MEROPS: M12.235; -;
 CC Genew: HGNC:13201; ADAMTS10.
 CC InterPro: IPR001762; Disintegrin.
 CC InterPro: IPR002870; Pep_M12B_propep.
 CC InterPro: IPR001590; Reprolysin.
 CC InterPro: IPR000884; TSPI.
 CC InterPro: IPR000130; Zn_M12peptidase.
 CC Pfam: PF00090; tsp_1; 5.
 CC Pfam: PF01421; Reprolysin; 1.
 CC Pfam: PF01562; Pep_M12B_propep; 1.
 CC SMART: SM00209; TSPI; 5.
 CC PROSITE: PS00215; ADAM_MEPRO; 1.
 CC PROSITE: PS00092; TSPI; 2.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 CC PROSITE: PS00427; DISINTEGRIN_1; FALSE NEG.
 CC Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
 CC Repeat; Extracellular matrix.
 CC REPEAT 1 1
 CC NON_TER <1 207 BY SIMILARITY.
 CC PROPEP 208 1077 ADAMTS-10.
 CC CHAIN

```

FT METAL 366 366 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 367 367 BY SIMILARITY.
FT METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 376 376 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 434 520 DISINTEGRIN-LIKE.
FT DOMAIN 578 679 CYS-RICH.
FT DOMAIN 680 802 SPACER.
FT DOMAIN 521 577 TSP TYPE-1 1.
FT DOMAIN 799 860 TSP TYPE-1 2.
FT DOMAIN 862 918 TSP TYPE-1 3.
FT DOMAIN 922 976 TSP TYPE-1 4.
FT DOMAIN 981 1031 TSP TYPE-1 5.
FT CARBOHYD 64 64 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 714 714 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 769 769 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 866 866 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1077 AA; 3914DE18DCBFF587 CRC64;

Query Match 24.0%; Score 172; DB 1; Length 1077;
Best Local Similarity 31.4%; Pred. No. 1.1e-09;
Matches 43; Conservative 24; Mismatches 56; Indels 14; Gaps 3;

QY 11 QEDTERVLTNLNIGAEELRLDPSLGAQFRVHLVKNVILTEPEGAPNITANLTSLLSVCG 70
DB 231 RRDVEQVLAIMVIAKLFODSLGSGVNLVTRLLILLTDEQPTLEITHAGKSLDSFCK 290

QY 71 WSOIT-----NPEDTDFCHADLVLYITRFDL-----ELPDGNRQVRGVTLGGACSP 119
DB 291 WKSIVNHSNGHNAIPENGVAHDTAVLTRIDYICIVKNKPCG---TLGLAPVGMGCERE 347

QY 120 WSLCLITDGTGFDLGVTI 136
DB 348 RSCSVNEDIGLPQAF 364

RESULT 10
AT58_HUMAN STANDARD; PRT; 890 AA.
ID AT58_HUMAN STANDARD; PRT; 890 AA.
AC Q9UP79; Q9NZS0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2)
DE (METH-8).
DE GN ADAMTS8 OR METH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=Lung;
RX MEDLINE=99367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
RA Lombard M., Iruela-Arispe M.L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
RT family of proteins with angio-inhibitory activity.";
RL J. Biol. Chem. 274:23349-23357(1999).
[2]
SEQUENCE OF 195-440 FROM N.A.
RX MEDLINE=20079168; PubMed=10610729;
RX Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS8, a novel metalloproteinase of the ADAM-TS family located on
RT mouse chromosome 9 and human chromosome 11.";
RL Genomics 62:312-315(1999).
CC -1- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER

```

```

CC EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND
CC KIDNEY.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTH: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF060153; AAD48081.1; -
CC HSSP; P34179; 1IAG.
CC MEROPS; M12.226; -
CC Genew; HGNC:224; ADAMTS8.
CC MIM; 605175; -
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; Zn_M12peptdse.
CC Pfam; PF00090; tsp_1; 2.
CC Pfam; PF01421; Reprolysin; 1.
CC SMART; SM00562; Pep_M12B_propep; 1.
CC SMART; SM00209; TSP1; 2.
CC PROSITE; PS00215; ADAM_MEPRO; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS00092; TSP1; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
CC Repeat; Extracellular matrix; Heparin-binding.
CC SIGNAL 1 27
CC PROPEP 28 214 BY SIMILARITY.
CC CHAIN 215 890 ADAMTS-8.
CC METAL 364 364 ZINC (CATALYTIC) (BY SIMILARITY).
CC ACT SITE 365 365 BY SIMILARITY.
CC METAL 368 368 ZINC (CATALYTIC) (BY SIMILARITY).
CC METAL 374 374 ZINC (CATALYTIC) (BY SIMILARITY).
CC DOMAIN 439 526 DISINTEGRIN-LIKE.
CC DOMAIN 527 583 TSP TYPE-1 1.
CC DOMAIN 584 690 CYS-RICH.
CC DOMAIN 691 832 SPACER.
CC DOMAIN 833 890 TSP TYPE-1 2.
CC DOMAIN 202 205 POLY-PRO.
CC CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 401 401 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 491 491 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 600 600 E -> R (IN REF. 2).
CC CONFLICT 195 195 YTELDDGHHGCLLDGCAALPLPTGL -> FSGCHLQGW
CC CONFLICT 413 410 IHFKYLCRCVSEKQDLMP (IN REF. 2).
CC SEQUENCE 890 AA; 96671 MW; 57D70EE03D5739D3 CRC64;

Query Match 23.7%; Score 170; DB 1; Length 890;
Best Local Similarity 26.6%; Pred. No. 1.4e-09;
Matches 33; Conservative 28; Mismatches 63; Indels 0; Gaps 0;

QY 13 DTERVLTNLNIGAEELRLDPSLGAQFRVHLVKNVILTEPEGAPNITANLTSLLSVCGWS 72
DB 239 DLQNHILTMVAARIYKHPISKNINLWVVKVLIVDEDEKMGVSDNGGLTLRNFNCNQ 298

QY 73 QTINPEDDTPGHADLVLYITRFDLPELPGNRQVRGVTLGGACSPWSLCLITDGTGFDL 132
DB 299 RRENQPSDRHPEHYDTAILLTRQNFQCGEGCLDTLGVADIGTICDPNKSQSVIEDEGLQA 358

```


CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
 CC COLLAGEN TYPE XIV.
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -!- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVEL
 CC IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS
 CC AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS
 CC WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -!- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF DERMATOSPARAXIS, A
 CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN
 CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN
 CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
 CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: X96389; CAA65253.1; -;
 CC DR MEROPS: M12.301; -;
 CC DR InterPro: IPR001762; Disintegrin.
 CC DR InterPro: IPR002870; Pep_M12B_propep.
 CC DR InterPro: IPR001590; RepPolysin.
 CC DR InterPro: IPR000884; TSP1.
 CC DR InterPro: IPR000130; Zn_MTpeptdase.
 CC DR Pfam: PF00090; tsp_1; 4_
 CC DR Pfam: PF01421; RepPolysin; 1.
 CC DR Pfam: PF01562; Pep_M12B_propep; 1.
 CC DR SMART: SM00209; TSP1; 4.
 CC DR PROSITE: PS00215; ADAM_MEPRO; 1.
 CC DR PROSITE: PS00092; TSP1; 1.
 CC DR PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.
 CC DR PROSITE: PS00427; DISINTEGRIN; 1.
 CC DR PROSITE: PS00427; DISINTEGRIN; 1.
 CC KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 CC Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.
 CC FT SIGNAL 1 28 POTENTIAL.
 CC FT PROPEP 29 253 BY SIMILARITY.
 CC FT CHAIN 254 1205 ADAMTS-2
 CC FT METAL 402 402 ZINC (CATALYTIC) (POTENTIAL).
 CC FT ACT_SITE 403 403 BY SIMILARITY.
 CC FT METAL 406 406 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT DOMAIN 474 554 DISINTEGRIN-LIKE.
 CC FT DOMAIN 555 611 TSP TYPE-1 1.
 CC FT DOMAIN 612 716 CYS-RICH.
 CC FT DOMAIN 717 845 SPACER.
 CC FT DOMAIN 846 905 TSP TYPE-1 2.
 CC FT DOMAIN 906 968 TSP TYPE-1 3.
 CC FT DOMAIN 969 1024 TSP TYPE-1 4.
 CC FT SITE 685 687 CELL ATTACHMENT SITE (POTENTIAL).
 CC FT DOMAIN 31 35 POLY-ALA.
 CC FT DOMAIN 177 180 POLY-GLU.
 CC FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 943 943 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 987 987 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1025 1025 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1092 1092 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1144 1144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1205 AA; 133887 MW; 7B5B232A45320371 CRC64;
 Query Match 22.2%; Score 159; DB 1; Length 1205;
 Best Local Similarity 33.3%; Pred. No. 2.7e-08;
 Matches 44; Conservative 23; Mismatches 57; Indels 8; Gaps 5;
 QY 2 VGPDIYQAH-QEDTERTYVLTNINIGALLRDPGLAQFRVHLVQVILTEPGAPNI-TA 59
 DB 268 VDDSVVQFHGTETVOKYLLTLNINIVNEIYHDSLGAIHNVILVRILLSYGKSMSTIETG 327
 QY 60 NLTSSLLSVCGRWSQ-TINPEDDTDPCHADLVLYITRFDLELDPGNRQVRGVLTOLGGACSP 118
 DB 328 NPSQSLSEYCRWAYLQQRDPDTHDEYH-DHAIFLTRQDF-----GPSGMOGYAPVTGWCHP 382
 QY 119 TWSCLITEDTGF 130
 DB 383 VRSCTLNHEDGF 394
 RESULT 13
 ATSS2 HUMAN
 ID ATSS2 HUMAN STANDARD; PRT; 1211 AA.
 AC Q95450;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
 DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS2) (ADAM-TS2)
 DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
 DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI)
 DE (Procollagen I/II amino-propeptide processing enzyme).
 GN ADAMTS2 OR PCINP OR PCPNI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI), AND DISEASE.
 RC TISSUE=Skin;
 RX MEDLINE=99347935; PubMed=10417273;
 RA Collige A., Sieron A.B., Li S.-W., Schwarze U., Petty E.,
 RA Wertelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,
 RA Byers P.H., Lapiere C.M., Prockop D.J., Nussgens B.V.;
 RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis
 RT are caused by mutations in the procollagen I N-proteinase gene.";
 RL Am. J. Hum. Genet. 65:308-317(1999).
 CC -!- FUNCTION: Cleaves the propeptides of type I and II collagen prior
 CC to fibril assembly. Does not act on type III collagen. May also
 CC play a role in development that is independent of its role in
 CC collagen biosynthesis.
 CC -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
 CC alpha-1(I) at Pro-[Gln and of alpha-1(II) and alpha-2(I) chains
 CC at Ala-[Gln.
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
 CC COLLAGEN TYPE XIV (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-
 CC PROCOLLAGEN PEPTIDASE ACTIVITY.
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON
 CC AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -!- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos
 CC syndrome type VIIC (EDS-VIIC), a recessively inherited connective-
 CC tissue disorder characterized clinically by severe skin fragility
 CC and joint hypermobility and biochemically by the presence in skin
 CC of procollagen incompletely processed at the N-terminus.

CC -/- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -/- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -/- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -/- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; AJ003125; CAA05880.1; -
CC DR MEROPS; M12.301; -
CC DR Genew; HGNC:218; ADAMTS2.
CC DR MIM; 604539; -
CC DR MIM; 225410; -
CC DR InterPro; IPR001762; Disintegrin.
CC DR InterPro; IPR002870; Pep_M12B_propep.
CC DR InterPro; IPR001590; Reptolysin.
CC DR InterPro; IPR000884; TSP1.
CC DR InterPro; IPR000130; Zn_Mtpeptdse.
CC DR Pfam; PF00090; tsp.1; 4.
CC DR Pfam; PF01421; Reptolysin.1.
CC DR Pfam; PF01562; Pep_M12B_propep; 1.
CC DR SMART; SMO0209; TSP1; 4.
CC DR PROSITE; PS0215; ADAM_MEPPO; 1.
CC DR PROSITE; PS0092; TSP1; 1.
CC DR PROSITE; PS00142; ZINC_PROTEASE; FALSE NEG.
CC DR PROSITE; PS00427; DISINTEGRIN.1; FALSE NEG.
CC DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
CC Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;
CC Alternative splicing; Ehlers-Danlos syndrome.
CC SIGNAL 1 29 POTENTIAL.
CC FT PROPEP 30 253 BY SIMILARITY.
CC FT CHAIN 254 1211 ADAMTS-2.
CC FT METAL 408 409 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT ACT_SITE 409 409 BY SIMILARITY.
CC FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT DOMAIN 480 560 DISINTEGRIN-LIKE.
CC FT DOMAIN 561 617 TSP TYPE-1 1.
CC FT DOMAIN 618 722 CYS-RICH.
CC FT SITE 691 693 CELL ATTACHMENT SITE (POTENTIAL).
CC FT DOMAIN 723 851 SPACER.
CC FT DOMAIN 852 911 TSP TYPE-1 2.
CC FT DOMAIN 912 974 TSP TYPE-1 3.
CC FT DOMAIN 975 1030 TSP TYPE-1 4.
CC FT DOMAIN 40 43 POLY-ALA.
CC FT DOMAIN 185 188 POLY-GLU.
CC FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 993 993 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 1031 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 1098 1098 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 1145 1145 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 1150 1150 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT VARSPLIC 544 566 HCFKGHCILPTDILKRDGSGA -> FRPGAVAHACVPS
CC FT VARSPLIC 567 1211 TLGGGRWIA (IN ISOFORM SPNP1).
CC FT VARSPLIC 1211 AA; 134722 MW; BECEE25C3CAD2D CRC64;
CC SEQUENCE 1211 AA; 134722 MW; BECEE25C3CAD2D CRC64;
CC -----
CC Query Match 21.9%; Score 157; DB 1; Length 1211;
CC Best Local Similarity 31.3%; Pred. No. 4.4e-08;
CC Matches 41; Conservative 24; Mismatches 60; Indels 6; Gaps 3;
CC -----
CC Qy 2 VGPVVFQAH-QEDTERTYVLTNIGAEILLRDPSSLAQAFRVHLVQVILTEPEGAPNI-TA 59
CC Db 274 VDDSVVQPHGKEHVQKYLTLNINVEIYHDESLGAHINVLVRLILSYGKMSLIBIG 333
CC Qy 60 NLTSSLLSVCVGSQINFEEDTDPGHADLVLYITRFDLELPDGNRQVRGVTLGGACSP 119

Db 334 NPSQSLNVCWAYLQKQKPTDGTGDEYHDAIFLTRQDF-----GPSGNGQYAPVTGMCHPV 389
Qy 120 WSLCLITDGTG 130
Db 390 RSCTLNHEDGF 400
RESULT 14
AT33_HUMAN STANDARD; PRT: 1205 AA.
ID AT33_HUMAN
AC O15072; Q9BXZ8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-3 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 3) (ADAM-TS 3) (Procollagen II
DE amino-proteptide processing enzyme) (Procollagen II N-proteinase) (PC
DE II-NP).
GN ADAMTS3 OR KIAA0366.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
-RN [1]
RP SEQUENCE OF 1-227 FROM N.A.
RX MEDLINE=21402912; PubMed=11408482;
RA Fernandes R.J., Hirohata S., Engle J.M., Colige A., Cohn D.H.,
RA Eyre D.R., Apte S.S.;
RT "Procollagen II amino propeptide processing by ADAMTS-3. Insights on
RT dermatoparaxis.";
RL J. Biol. Chem. 276:31502-31509(2001).
RN [2]
RP SEQUENCE OF 5-1205 FROM N.A.
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RL code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
CC -/- FUNCTION: Cleaves the propeptides of type II collagen prior to
CC fibril assembly. Does not act on types I and III collagens.
CC -/- COPACITOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -/- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (BY similarity).
CC -/- TISSUE SPECIFICITY: Found in cartilage and skin.
CC -/- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -/- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -/- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -/- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -/- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -/- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; AF247668; AAK28400.1; -
CC DR EMBL; AB002364; BAA20821.1; -
CC DR MEROPS; M12.220; -
CC DR Genew; HGNC:219; ADAMTS3.
CC DR MIM; 605011; -
CC DR InterPro; IPR001762; Disintegrin.
CC DR InterPro; IPR002870; Pep_M12B_propep.
CC DR InterPro; IPR001590; Reptolysin.


```

DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn MTpeptdae.
DR Pfam; PF00090; tspl_1; 4.
DR Pfam; PF01421; RepPolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPl; 4.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 20
FT PROPEP 21 249
FT CHAIN 250 1205
FT METAL 398 398
FT ACT_SITE 399 399
FT METAL 402 402
FT METAL 408 408
FT DOMAIN 470 550
FT DOMAIN 551 607
FT DOMAIN 608 712
FT DOMAIN 713 844
FT DOMAIN 845 902
FT DOMAIN 903 965
FT DOMAIN 966 1017
FT DOMAIN 1017 1017
FT CARBOHYD 246 249
FT CARBOHYD 83 83
FT CARBOHYD 119 119
FT CARBOHYD 242 242
FT CARBOHYD 345 345
FT CARBOHYD 475 475
FT CARBOHYD 814 814
FT CARBOHYD 942 942
FT SEQUENCE 1205 AA; 135570 MW; EB07B286FC95FB7 CRC64;

Query Match 21.4%; Score 153; DB 1; Length 1205;
Best Local Similarity 31.3%; Pred. No. 1.1e-07;
Matches 41; Conservative 21; Mismatches 63; Indels 6; Gaps 3;

QY 2 VQPDVQAH-OEDTERYVLTNLNIGALLRDPGLQAFVHLVQWVILTEPGAPNI-TA 59
Db 264 VQDSVVRFGKHEVQVNYLLTLMNIVNEIYHDESGLVHINVLVRLMGLVAKSISLIERG 323
QY 60 NLTSSLLSCVGSQNTINPEDDTPGHADLVLTITFDELPGDNROVRGVGTOLGGACSP 119
Db 324 NPSRSLENCVRWASQQQRSDLNHSEHHDHAFILTRQDF---GPAGMQGVAPVTGMCHPV 379
QY 120 WSLITEDTGF 130
Db 380 RSLTLNHEDGF 390

RESULT 15
AT55.MOUSE
ID AT55.MOUSE STANDARD; PRT; 930 AA.
AC Q9R061.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motif 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
DE (ADMP-2) (Implantin).
GN ADAMTS5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99395124; PubMed=10464288;
RA Hurekainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of

```

```

zinc metalloproteases.";
J. Biol. Chem. 274:25555-25563(1999).
-!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
-!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393
site.
-!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
-!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
-!- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PERI-
IMPLANTATION PERIOD IN EMBRYO AND TROPHOBLAST AND AT LOW OR
UNDETECTABLE LEVEL THEREAFTER.
-!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
-!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
SIMILARITY).
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
-!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
-!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; AF140673; AAD56356.1; -.
CC MEROPS; M12.225; -.
CC MGD; MGI:1346321; Adamts5.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; RepPolysin.
CC InterPro; IPR000884; TSPl.
CC InterPro; IPR000130; Zn MTpeptdae.
CC Pfam; PF00090; tspl_1; 2.
CC Pfam; PF01421; RepPolysin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC SMART; SM00209; TSPl; 2.
CC PROSITE; PS50215; ADAM_MEPRO; 1.
CC PROSITE; PS00092; TSPl; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 21
FT PROPEP 22 261
FT CHAIN 262 930
FT SITE 209 209
FT SITE 410 410
FT METAL 411 411
FT ACT_SITE 414 414
FT METAL 420 420
FT METAL 485 566
FT DOMAIN 567 623
FT DOMAIN 624 731
FT DOMAIN 732 874
FT DOMAIN 875 930
FT DOMAIN 930 930
FT DOMAIN 930 930
FT DOMAIN 930 930
FT CARBOHYD 498 498
FT CARBOHYD 728 728
FT CARBOHYD 802 802
FT CARBOHYD 807 807
FT SEQUENCE 930 AA; 101780 MW; 84DE84B26170D4DC CRC64;

Query Match 20.7%; Score 148.5; DB 1; Length 930;
Best Local Similarity 28.7%; Pred. No. 2.4e-07;
Matches 35; Conservative 22; Mismatches 64; Indels 1; Gaps 1;

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:34:44 ; Search time 22.0541 Seconds
(without alignments)
329.684 Million cell updates/sec

Title: SEQID_15
Perfect score: 716
Sequence: 1 AVGPVDFQAHQEDTERVYLT.....SPTWSCLITDGFGLGVTI 136

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 221153

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW PUB.pap.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW PUB.pap.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US03_NEW PUB.pap.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW PUB.pap.*
 - 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
 - 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW PUB.pap.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pap.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW PUB.pap.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pap.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW PUB.pap.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	716	100.0	136	10	US-09-833-328-15
2	716	100.0	148	10	US-09-833-328-4
3	716	100.0	150	10	US-09-833-328-6
4	716	100.0	242	10	US-09-978-979-1
5	716	100.0	1416	10	US-09-836-712-2
6	713	99.6	203	10	US-09-781-080B-2
7	713	99.6	1120	10	US-09-781-080B-11
8	702	98.0	133	10	US-09-833-328-2
9	212.5	29.7	997	10	US-09-918-171A-7
10	206.5	28.8	1690	10	US-09-788-043C-5
11	204.5	28.6	730	9	US-10-163-316-2
12	201	28.1	2150	10	US-09-321-987B-2
13	201	28.1	2165	10	US-09-800-729-155
14	187.5	26.2	947	10	US-09-788-043C-7
15	187.5	26.2	1629	10	US-09-972-467-2
16	184	25.7	1882	10	US-09-918-171A-13
17	180.5	25.2	839	9	US-09-912-788-2
18	178.5	24.9	727	9	US-10-097-597-1
19	178.5	24.9	727	9	US-10-097-580-1

20	178.5	24.9	727	10	US-09-445-023A-1	Sequence 1, Appli
21	178.5	24.9	967	12	US-10-105-929-2	Sequence 2, Appli
22	177.5	24.8	727	9	US-10-097-597-12	Sequence 12, Appli
23	177.5	24.8	727	9	US-10-097-580-12	Sequence 12, Appli
24	177.5	24.8	727	10	US-09-445-023B-12	Sequence 12, Appli
25	177.5	24.8	950	10	US-09-321-987B-4	Sequence 4, Appli
26	177.5	24.8	968	9	US-10-163-316-7	Sequence 7, Appli
27	176	24.6	317	10	US-09-963-791-16	Sequence 16, Appli
28	176	24.6	356	10	US-09-963-791-20	Sequence 20, Appli
29	176	24.6	438	10	US-09-963-791-22	Sequence 22, Appli
30	176	24.6	468	10	US-09-963-791-6	Sequence 6, Appli
31	176	24.6	507	10	US-09-963-791-10	Sequence 10, Appli
32	176	24.6	589	10	US-09-963-791-12	Sequence 12, Appli
33	176	24.6	757	10	US-09-963-791-24	Sequence 24, Appli
34	176	24.6	859	10	US-09-918-171A-5	Sequence 5, Appli
35	176	24.6	908	10	US-09-963-791-2	Sequence 2, Appli
36	176	24.6	959	10	US-09-788-043C-1	Sequence 1, Appli
37	173.5	24.2	823	9	US-10-163-316-2	Sequence 2, Appli
38	173.5	24.2	823	9	US-09-965-631-4	Sequence 4, Appli
39	173.5	24.2	814	10	US-09-858-081-12	Sequence 12, Appli
40	172	24.0	947	10	US-09-788-043C-3	Sequence 3, Appli
41	172	24.0	1044	10	US-09-858-081-9	Sequence 9, Appli
42	172	24.0	1068	10	US-09-858-081-2	Sequence 2, Appli
43	172	24.0	1133	10	US-09-858-068-2	Sequence 2, Appli
44	172	24.0	1133	10	US-09-858-068-2	Sequence 2, Appli
45	170	23.7	245	10	US-09-918-171A-11	Sequence 11, Appli

ALIGNMENTS

RESULT 1
US-09-833-328-15
; Sequence 15, Application US/09833328
; Patent No. US2002036713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Scheiflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Rndolf
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Gericeen, Helena E.
; TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protea
; TITLE OF INVENTION: Comprising a Polypeptide Chain with the Amino Acid Sequenc
; FILE REFERENCE: 247.00CIP
; CURRENT APPLICATION NUMBER: US/09/833.328
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 15
; LENGTH: 136
; TYPE: PRT
; ORGANISM: human
US-09-833-328-15

Query Match 100.0%; Score 716; DB 10; Length 136;
Best Local Similarity 100.0%; Pred. No. 2e-73;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AVGPVDFQAHQEDTERVYLTNINIGALLRDPGLGAQFRVHLVQKVVILTEPGAPNITAN 60
Db 1 AVGPVDFQAHQEDTERVYLTNINIGALLRDPGLGAQFRVHLVQKVVILTEPGAPNITAN 60
Qy 61 LTSSLLSVCGWSQTNPEDDTPGHADLVLYITRFDLELPDGNRVRGVTOLGGACSPTW 120
Db 61 LTSSLLSVCGWSQTNPEDDTPGHADLVLYITRFDLELPDGNRVRGVTOLGGACSPTW 120

QY 121 SCLITDGTGFDLGVTI 136
Db 121 SCLITDGTGFDLGVTI 136

RESULT 2

US-09-833-328-4
; Sequence 4, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwartz, Hans-Peter
; APPLICANT: Scheiflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Randolph
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Geritsen, Helena E.
; TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
; FILE REFERENCE: 247.00CIP
; CURRENT APPLICATION NUMBER: US/09/833,328
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 148
; TYPE: PRT
; ORGANISM: human
US-09-833-328-4

Query Match 100.0%; Score 716; DB 10; Length 148;
Best Local Similarity 100.0%; Pred. No. 2.3e-73;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVGPDVFOAQHEDTERVYLTNLNIGAEALLRDPGLAQFRVHLVKMVLTEPEGAPNITAN 60
Db 13 AVGPDVFOAQHEDTERVYLTNLNIGAEALLRDPGLAQFRVHLVKMVLTEPEGAPNITAN 72
QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPGNRQVRGVTLGGACSPW 120
Db 73 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPGNRQVRGVTLGGACSPW 132
QY 121 SCLITDGTGFDLGVTI 136
Db 133 SCLITDGTGFDLGVTI 148

RESULT 3

US-09-833-328-6
; Sequence 6, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwartz, Hans-Peter
; APPLICANT: Scheiflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Randolph
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Geritsen, Helena E.
; TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
; FILE REFERENCE: 247.00CIP
; CURRENT APPLICATION NUMBER: US/09/833,328
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254

; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 150
; TYPE: PRT
; ORGANISM: human
US-09-833-328-6

Query Match 100.0%; Score 716; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.3e-73;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVGPDVFOAQHEDTERVYLTNLNIGAEALLRDPGLAQFRVHLVKMVLTEPEGAPNITAN 60
Db 15 AVGPDVFOAQHEDTERVYLTNLNIGAEALLRDPGLAQFRVHLVKMVLTEPEGAPNITAN 74
QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPGNRQVRGVTLGGACSPW 120
Db 75 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPGNRQVRGVTLGGACSPW 134
QY 121 SCLITDGTGFDLGVTI 136
Db 135 SCLITDGTGFDLGVTI 150

RESULT 4

US-09-978-979-1
; Sequence 1, Application US/09978979
; Patent No. US20020151702A1
; GENERAL INFORMATION:
; APPLICANT: Racie, Lisa, A.
; Agostino, Michael, J.
; Wolfman, Neil
; Morris, Elisabeth
; TITLE OF INVENTION: Aggrecanase Molecules
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Home Products Corporation
; STREET: One Campus Drive
; CITY: Parsippany
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/978,979
; FILING DATE: 16-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/60/241,469
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Kapinos, Ellen J.
; REGISTRATION NUMBER: 32,345
; REFERENCE/DOCKET NUMBER: GI 5435p
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (973) 660-5000
; TELEFAX: (973) 683-4117
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 242 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-978-979-1

Query Match 100.0%; Score 716; DB 10; Length 242;
Best Local Similarity 100.0%; Pred. No. 4.4e-73;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVGPVFOAQHQEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 60
DB 54 AVGPVFOAQHQEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 113

QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 120
DB 114 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 173

QY 121 SCLITEDTGFDLGVTI 136
DB 174 SCLITEDTGFDLGVTI 189

RESULT 5
US-09-836-712-2
; Sequence 2, Application US/09836712
; Patent No. US20010049106A1
; GENERAL INFORMATION:
; APPLICANT: PEIZER INC.
; TITLE OF INVENTION: ADAMTS POLYPEPTIDES, NUCLEIC ACIDS ENCODING THEM, AND
; FILE REFERENCE: PC10851A
; CURRENT APPLICATION NUMBER: US/09/836,712
; CURRENT FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1416
; TYPE: PRT
; ORGANISM: Human
US-09-836-712-2

Query Match 100.0%; Score 716; DB 10; Length 1416;
Best Local Similarity 100.0%; Pred. No. 4.6e-72;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVGPVFOAQHQEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 60
DB 110 AVGPVFOAQHQEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 169

QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 120
DB 170 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 229

QY 121 SCLITEDTGFDLGVTI 136
DB 230 SCLITEDTGFDLGVTI 245

RESULT 6
US-09-781-080B-2
; Sequence 2, Application US/09781080B
; Patent No. US20020142439A1
; GENERAL INFORMATION:
; APPLICANT: HOLLOWAY, JIM
; APPLICANT: SHEPPARD, PAUL
; TITLE OF INVENTION: Anti-Angiogenic Intestinal Peptides,
; FILE REFERENCE: zidnt5
; CURRENT APPLICATION NUMBER: US/09/781,080B
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-781-080B-2
Query Match 99.6%; Score 713; DB 10; Length 203;
Best Local Similarity 99.3%; Pred. No. 7.5e-73;
Matches 135; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVGPVFOAQHQEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 60
DB 14 AVGPVFOAQHQEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 73

QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 120
DB 74 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 133

QY 121 SCLITEDTGFDLGVTI 136
DB 134 SCLITEDTGFDLGVTI 149

RESULT 7
US-09-781-080B-11
; Sequence 11, Application US/09781080B
; Patent No. US20020142439A1
; GENERAL INFORMATION:
; APPLICANT: HOLLOWAY, JIM
; APPLICANT: SHEPPARD, PAUL
; APPLICANT: YAMAMOTO, GAYLE
; TITLE OF INVENTION: Anti-Angiogenic Intestinal Peptides,
; FILE REFERENCE: zidnt5
; CURRENT APPLICATION NUMBER: US/09/781,080B
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 1120
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1120)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-781-080B-11

Query Match 99.6%; Score 713; DB 10; Length 1120;
Best Local Similarity 99.3%; Pred. No. 7.4e-72;
Matches 135; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVGPVFOAQHQEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 60
DB 117 AVGPVFOAQHQEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 176

QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 120
DB 177 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 236

QY 121 SCLITEDTGFDLGVTI 136
DB 237 SCLITEDTGFDLGVTI 252

RESULT 8
US-09-833-328-2
; Sequence 2, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Scheiflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Rndolf
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus

```
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Gexitsen, Helena E.
; TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
; TITLE OF INVENTION: Comprising a Polypeptide Chain with the Amino Acid Sequence AAGG
; FILE REFERENCE: 247.00CIP
; CURRENT APPLICATION NUMBER: US/09/833,328
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 133
; TYPE: PRT
; ORGANISM: human
US-09-833-328-2

Query Match      98.0%; Score 702; DB 10; Length 133;
Best Local Similarity 100.0%; Pred. No. 7.5e-72;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PDVFOAHOEDTERVYLTNLNIGAEALLRDPISLGAQFRVHLVKMVLITEPEGAPNITANLTS 63
Db 1 PDVFOAHOEDTERVYLTNLNIGAEALLRDPISLGAQFRVHLVKMVLITEPEGAPNITANLTS 60

Qy 64 SLLSVCGWSQTINPEDDDTPGHADLVLYITRFDLELPGNRQVRGVTOGLGACSPFTWSCL 123
Db 61 SLLSVCGWSQTINPEDDDTPGHADLVLYITRFDLELPGNRQVRGVTOGLGACSPFTWSCL 120

Qy 124 ITEDTGFDLGVTI 136
Db 121 ITEDTGFDLGVTI 133

RESULT 9
US-09-918-171A-7
; Sequence 7, Application US/09918171A
; Patent No. US20020110894A1
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/04193
; CURRENT APPLICATION NUMBER: US/09/918,171A
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 09/369,364
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
US-09-918-171A-7

Query Match      29.7%; Score 212.5; DB 10; Length 997;
Best Local Similarity 37.5%; Pred. No. 2e-15;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;

Qy 11 QEDTERVYLTNLNIGAEALLRDPISLGAQFRVHLVKMVLITEPEGAPNITANLTSLLSVCG 70
Db 260 PQVESYVLTIMNVAGLFHDPISGNPIHITIVRLVLEDEEDLKITHADNTKSFCK 319

Qy 71 WSQTINPEDDDTPGHADLVLYITRFDLELPGNR--QVRGVTOGLGACSPFTWSCLITEDT 128
Db 320 WQKSINMGDAHPLHDDTAILLTRKDL-CAAMNRPCTGLSLHVGMCQPHRSCSINEDT 378

Qy 129 GFDLGVTI 136
Db 379 GLPLAFTV 386
```

```
RESULT 10
US-09-788-043C-5
; Sequence 5, Application US/09788043C
; Patent No. US20020107361A1
; GENERAL INFORMATION:
; APPLICANT: Heller, Renu
; APPLICANT: Zuo, Fengrong
; APPLICANT: Klonowski, Paul
; TITLE OF INVENTION: No. US20020107361A1e1 Metalloproteases Having
; TITLE OF INVENTION: Thrombospondin Domains and Nucleic Acid Compositions
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: ROCH-004
; CURRENT APPLICATION NUMBER: US/09/788,043C
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/184,152
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1690
; TYPE: PRT
; ORGANISM: human
US-09-788-043C-5

Query Match      28.8%; Score 206.5; DB 10; Length 1690;
Best Local Similarity 36.7%; Pred. No. 1.9e-14;
Matches 47; Conservative 22; Mismatches 56; Indels 3; Gaps 2;

Qy 11 QEDTERVYLTNLNIGAEALLRDPISLGAQFRVHLVKMVLITEPEGAPNITANLTSLLSVCG 70
Db 264 PQVESYVLTIMNVAGLFHDPISGNPIHITIVRLVLEDEEDLKITHADNTKSFCK 323

Qy 71 WSQTINPEDDDTPGHADLVLYITRFDLELPGNR--QVRGVTOGLGACSPFTWSCLITEDT 128
Db 324 WQKSINMGDAHPLHDDTAILLTRKDL-CATMNRPCETGLSLHVGMCQPHRSCSINEDT 382

Qy 129 GFDLGVTI 136
Db 383 GLPLAFTV 390

RESULT 11
US-10-163-547-16
; Sequence 16, Application US/10163547
; Publication No. US20030037350A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; APPLICANT: Myer, Rachel
; APPLICANT: Rudolf-Owen, Laura A.
; TITLE OF INVENTION: No. US20030037350A1e1 nucleic acid sequences encoding a
; TITLE OF INVENTION: human ubiquitin protease, lipase, dynamin, short chain
; TITLE OF INVENTION: dehydrogenase, and ADAM-TS Metalloprotease and uses
; FILE REFERENCE: 35800/247836(5800-243)
; CURRENT APPLICATION NUMBER: US/10/163,547
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 09/407,356
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 09/704,918
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 09/435,311
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 09/796,100
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,503
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: 09/781,598
; PRIOR FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/182,009
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 09/782,952
```

Patent No. US20020068319A1
GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: 32 Human secreted proteins
FILE REFERENCE: P204P1
CURRENT APPLICATION NUMBER: US/09/800,729
CURRENT FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: PCT/US00/26013
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: 60/155,709
PRIOR FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 217
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 155
LENGTH: 2165
TYPE: PRT
ORGANISM: Homo sapiens
US-09-800-729-155

Query Match 28.1%; Score 201; DB 10; Length 2165;
Best Local Similarity 33.6%; Pred. No. 1.1e-13;
Matches 44; Conservative 17; Mismatches 70; Indels 0; Gaps 0;
QY 6 VFQAHQEDTERRYVLTNLNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSL 65
DB 292 MYEYHGRSLEDYVLTFLSTVASIYRHQSLRASINVVVVKLVLTENAGPRITQNAOQTL 351
QY 66 LSVCGWSQTNIPEDDTPGHADLVLYITRFDLELPDGNRQVRGVLTOLGGACSPWWSCLIT 125
DB 352 QDFCRWQYYNDPDDSSVQHHDVAILLTRKDCIRSGKCDTLGLAEGLTMDMOKSCAII 411
QY 126 EDTGFDLGVTI 136
DB 412 EDNGLSAAFTI 422

RESULT 14
US-09-788-043C-7
Sequence 7, Application US/09788043C
Patent No. US20020107361A1
GENERAL INFORMATION:
APPLICANT: Heller, Renu
APPLICANT: Zuo, Fengrong
APPLICANT: Klonowski, Paul
TITLE OF INVENTION: No. US20020107361A1 Metalloproteases Having
TITLE OF INVENTION: Thrombospondin Domains and Nucleic Acid Compositions
TITLE OF INVENTION: Encoding the Same
FILE REFERENCE: ROCH-004
CURRENT APPLICATION NUMBER: US/09/788,043C
CURRENT FILING DATE: 2001-02-16
PRIOR APPLICATION NUMBER: 60/184,152
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 947
TYPE: PRT
ORGANISM: human
US-09-788-043C-7

Query Match 26.2%; Score 187.5; DB 10; Length 947;
Best Local Similarity 30.2%; Pred. No. 1.2e-12;
Matches 39; Conservative 28; Mismatches 55; Indels 7; Gaps 2;
QY 10 HQEDTERRYVLTNLNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSL 69
DB 309 HGENLQHYILTLMSIVASIVKPSIGNLVINIVIVLNIVHNEQDGPISFNAGTTLKNFC 368
QY 70 GWSQTNIPEDDTPG--HADLVLYITRFDLELPDGNRQVRGVLTOLGGACSPWWSCLIT 127
DB 369 QWQHSKN-----SPGGIHHDTAVLLTRQDICRAHDKCDTLGLAEGLTICDPYRSCSISED 423
QY 128 TGFDLGVTI 136

Query Match 28.1%; Score 204.5; DB 9; Length 730;
Best Local Similarity 34.6%; Pred. No. 1e-14;
Matches 44; Conservative 26; Mismatches 54; Indels 3; Gaps 2;
QY 12 EDTERYVLTNLNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSLVCGW 71
DB 265 ENVESYIITIMVVTGLPHNPSIGNAIIHVVRLLLEEEEOGLKIVHHAETLSSFCCKW 324
QY 72 SQTNIPEDDTPGHADLVLYITRFDLELPDGNR--QVRGVLTOLGGACSPWWSCLIT 129
DB 375 QKSINPKSDLPVHHGVAVLLTRKDI-CAGFNRPCTETLGLSHLGMCPHRSNINEDSG 383
QY 130 FDLGVTI 136
DB 384 LPLAFTI 390

PRIOR FILING DATE: 2001-02-14
PRIOR APPLICATION NUMBER: 60/182,408
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: 09/496,005
PRIOR FILING DATE: 2000-02-01
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 730
TYPE: PRT
ORGANISM: Homo sapiens
US-10-163-547-16

Query Match 28.1%; Score 201; DB 10; Length 2150;
Best Local Similarity 33.6%; Pred. No. 1.1e-13;
Matches 44; Conservative 17; Mismatches 70; Indels 0; Gaps 0;
QY 6 VFQAHQEDTERRYVLTNLNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSL 65
DB 292 MYEYHGRSLEDYVLTFLSTVASIYRHQSLRASINVVVVKLVLTENAGPRITQNAOQTL 351
QY 66 LSVCGWSQTNIPEDDTPGHADLVLYITRFDLELPDGNRQVRGVLTOLGGACSPWWSCLIT 125
DB 352 QDFCRWQYYNDPDDSSVQHHDVAILLTRKDCIRSGKCDTLGLAEGLTMDMOKSCAII 411
QY 126 EDTGFDLGVTI 136
DB 412 EDNGLSAAFTI 422

RESULT 13
US-09-800-729-155
Sequence 155, Application US/09800729
Patent No. US20020102210A1
GENERAL INFORMATION:
APPLICANT: Kimble, Judith E
APPLICANT: Bleiloch, Robert H
TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
FILE REFERENCE: 960296.95386
CURRENT APPLICATION NUMBER: US/09/321,987B
CURRENT FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: 60/087,170
PRIOR FILING DATE: 1998-05-29
PRIOR APPLICATION NUMBER: 60/129,023
PRIOR FILING DATE: 1999-04-13
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 2150
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-09-321-987B-2

Query Match 28.1%; Score 201; DB 10; Length 2150;
Best Local Similarity 33.6%; Pred. No. 1.1e-13;
Matches 44; Conservative 17; Mismatches 70; Indels 0; Gaps 0;
QY 6 VFQAHQEDTERRYVLTNLNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSL 65
DB 292 MYEYHGRSLEDYVLTFLSTVASIYRHQSLRASINVVVVKLVLTENAGPRITQNAOQTL 351
QY 66 LSVCGWSQTNIPEDDTPGHADLVLYITRFDLELPDGNRQVRGVLTOLGGACSPWWSCLIT 125
DB 352 QDFCRWQYYNDPDDSSVQHHDVAILLTRKDCIRSGKCDTLGLAEGLTMDMOKSCAII 411
QY 126 EDTGFDLGVTI 136
DB 412 EDNGLSAAFTI 422

Query Match 28.1%; Score 201; DB 10; Length 2150;
Best Local Similarity 33.6%; Pred. No. 1.1e-13;
Matches 44; Conservative 17; Mismatches 70; Indels 0; Gaps 0;
QY 6 VFQAHQEDTERRYVLTNLNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSL 65
DB 292 MYEYHGRSLEDYVLTFLSTVASIYRHQSLRASINVVVVKLVLTENAGPRITQNAOQTL 351
QY 66 LSVCGWSQTNIPEDDTPGHADLVLYITRFDLELPDGNRQVRGVLTOLGGACSPWWSCLIT 125
DB 352 QDFCRWQYYNDPDDSSVQHHDVAILLTRKDCIRSGKCDTLGLAEGLTMDMOKSCAII 411
QY 126 EDTGFDLGVTI 136
DB 412 EDNGLSAAFTI 422

Query Match 28.1%; Score 201; DB 10; Length 2150;
Best Local Similarity 33.6%; Pred. No. 1.1e-13;
Matches 44; Conservative 17; Mismatches 70; Indels 0; Gaps 0;
QY 6 VFQAHQEDTERRYVLTNLNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSL 65
DB 292 MYEYHGRSLEDYVLTFLSTVASIYRHQSLRASINVVVVKLVLTENAGPRITQNAOQTL 351
QY 66 LSVCGWSQTNIPEDDTPGHADLVLYITRFDLELPDGNRQVRGVLTOLGGACSPWWSCLIT 125
DB 352 QDFCRWQYYNDPDDSSVQHHDVAILLTRKDCIRSGKCDTLGLAEGLTMDMOKSCAII 411
QY 126 EDTGFDLGVTI 136
DB 412 EDNGLSAAFTI 422

Db 424 SGLSTAFTI 432

RESULT 15

US-09-972-467-2
; Sequence 2, Application US/09972467
; Patent No. US20020090373A1
; GENERAL INFORMATION:
; APPLICANT: PFIZER INC.
; TITLE OF INVENTION: ADAMTS POLYPEPTIDES, NUCLEIC ACIDS ENCODING THEM, AND
; FILE OF INVENTION: USES THEREOF
; FILE REFERENCE: PC10850A
; CURRENT APPLICATION NUMBER: US/09/972.467
; CURRENT FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1629
; TYPE: PRT
; ORGANISM: Human
US-09-972-467-2

Query Match 26.2%; Score 187.5; DB 10; Length 1629;
Best Local Similarity 30.2%; Pred. No. 2.5e-12;
Matches 39; Conservative 28; Mismatches 55; Indels 7; Gaps 2;

QY 10 HQEDTERYVLTNLNIGAEELLRDPGLGAQFRVHLVKMVLTEPEGAPNITANLTSSLLSVC 69
Db 309 HGENLOHYILTMSIVASYKDFSGINLVIVNLVHNEODGPSISFNAOTTLKNFC 368
QY 70 GWSQTINPEDDTPG--HADLVLYITRFDLEPDGNGRQVRGVTQLGACGCSPTWSCLITED 127
Db 369 QMOHKN-----SPGGIHHDVALLTRQDICRAHDKCDTLGLAELGTICDPYRSCSISED 423
QY 128 TGFDLGVTI 136
Db 424 SGLSTAFTI 432

Search completed: March 20, 2003, 12:36:38
Job time : 26.0541 sec

GenCore version 5.1.4.p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:32:34 ; Search time 2.18919 Seconds
(without alignments)
526.959 Million cell updates/sec

Title: SEQID_1
Perfect score: 57
Sequence: 1 AAGGILHLELLV 12
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:
1: pir1:
2: pir2:
3: pir3:
4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	40	70.2	489	1 S36390	mitochondrial proc
2	38	66.7	316	2 T08807	mitochondrial proc
3	38	66.7	514	2 A31643	cell adhesion 80K
4	38	66.7	514	2 A44100	cell adhesion mole
5	38	66.7	526	2 S43561	YCR37C homolog K08
6	38	66.7	534	2 G88575	protein K08E5.2 li
7	38	66.7	1081	2 T31094	surface antigen B8
8	37	64.9	116	2 A72502	hypothetical prote
9	37	64.9	356	2 B90387	hypothetical prote
10	37	64.9	378	2 S12190	mobl protein - Thi
11	37	64.9	876	2 D70971	hypothetical prote
12	36	63.2	233	2 T35594	hypothetical prote
13	36	63.2	392	2 A44167	aminomethyltransf
14	36	63.2	480	2 A48043	ubiquinol-cytochr
15	36	63.2	505	2 C98300	hypothetical prote
16	36	63.2	505	2 AD2983	hypothetical prote
17	36	63.2	1096	2 A26879	pullulanase (EC 3.
18	36	63.2	1142	2 S62714	phyochrome lb - M
19	36	63.2	1178	2 S57535	probable membrane
20	36	63.2	1829	2 T14280	RW1 protein - mous
21	35	61.4	145	1 UOCA1	urotensin I precu
22	35	61.4	165	2 G90142	hypothetical prote
23	35	61.4	189	2 A32840	hypothetical 22K p
24	35	61.4	305	2 T15710	hypothetical prote
25	35	61.4	362	1 ZP80C1	ubiquinol-cytochr
26	35	61.4	385	2 AH2200	two-component hydr
27	35	61.4	458	2 H82200	sodium/alanine sym
28	35	61.4	691	2 S75863	translation elonga
29	35	61.4	692	2 AC2348	translation elonga

30	35	61.4	694	1 S04429	translation elonga
31	35	61.4	697	2 S04390	translation elonga
32	35	61.4	698	2 E75536	translation elonga
33	35	61.4	805	2 T48840	transducer protein
34	35	61.4	921	2 D86293	PH2-22 protein -
35	35	61.4	1038	2 S38801	pullulanase (EC 3.
36	35	61.4	1090	2 S11823	pullulanase (EC 3.
37	34	59.6	204	2 A71368	hypothetical prote
38	34	59.6	214	2 G83882	hypothetical prote
39	34	59.6	359	2 A22891	fix2 protein - Rhi
40	34	59.6	360	2 F96902	aspartate-semialde
41	34	59.6	382	2 T20655	hypothetical prote
42	34	59.6	399	2 AE3556	alanine racemase (
43	34	59.6	401	2 A75388	NADH dehydrogenase
44	34	59.6	403	2 D87597	hypothetical prote
45	34	59.6	424	2 T35924	hypothetical prote

ALIGNMENTS

RESULT 1

S36390 Mitochondrial processing peptidase (EC 3.4.24.64) beta chain precursor [validated]
N:Alternate names: beta-MPP; mitochondrial processing peptidase chain P52
C:Species: Rattus norvegicus (Norway rat)
C:Date: 09-Dec-1991 #sequence revision 14-Jul-1994 #text_change 31-Dec-2000
C:Accession: S36390; S36391; F01229
R:Pages, V.; Rosenberg, L.E.; Fenton, W.A.; Kalousek, F.
Proc. Natl. Acad. Sci. U.S.A. 90, 5355-5358, 1993
A:Title: The beta subunit of the mitochondrial processing peptidase from rat liver
A:Reference number: S36390; MUID:93281757; PMID:8506385
A:Accession: S36390
A:Molecule type: mRNA
A:Residues: 1-489 <PAC>
A:Cross-references: GB:L12965; NID:G294588; PIDN:AAA41633.1; PID:G294589
A:Accession: S36391
A:Molecule type: Protein
A:Residues: 46-67; 92-104; 116-125; 409-423; 432-441 <PA2>
C:Genetics:
A:Genome: nuclear
C:Superfamily: mitochondrial processing peptidase alpha chain
C:Keywords: heterodimer; hydrolase; metalloproteinase; mitochondrial matrix; mitoc
F:1-45/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F:46-489/Product: mitochondrial processing peptidase beta chain #status experiment
F:101,105/Binding site: zinc (His) #status predicted
F:104/Active site: Glu #status predicted

Query Match 70.2%; Score 40; DB 1; Length 489;
Best Local Similarity 72.7%; Pred. No. 9.7;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGILHLELL 11
||||: |||
Db 243 AAGGVCHNELL 253

RESULT 2

T08807 Mitochondrial processing peptidase (EC 3.4.24.64) beta chain (similarity) - human
N:Alternate names: protein DKF2p586i1223.1
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Feb-2001
C:Accession: T08807
R:Ansorge, W.; Winkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, May 1999
A:Reference number: 216472
A:Accession: T08807
A:Molecule type: mRNA
A:Residues: 1-316 <ANS>
A:Cross-references: EMBL:AL050216
A:Experimental source: adult uterus; clone DKF2p586i1223
C:Genetics:

A>Note: DXFZ58611223.1
 C:Superfamily: mitochondrial processing peptidase alpha chain
 C:Keywords: hydrolase; metalloproteinase; mitochondrion; oxidative phosphorylation; oxid

Query Match 66.7%; Score 38; DB 2; Length 316;
 Best Local Similarity 72.7%; Pred. No. 15;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGILHLELL 11
 |||||
 DB 70 AAGGVSHDELL 80

RESULT 3
 A31643
 cell adhesion 80K glycoprotein precursor - slime mold (Dictyostelium discoideum)
 C:Alternate names: cell adhesion molecule csa; cell surface glycoprotein, 80K
 C:Species: Dictyostelium discoideum
 A:Variety: strain AX-2
 C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 01-Dec-2000
 C:Accession: S22066; A31643; A26310; A23951; A33463; S01473
 R:Desbarats, L.; Lam, T.Y.; Wong, L.M.; Siu, C.H.
 submitted to the EMBL Data Library, June 1992
 A:Description: Identification of a unique camp-responsive element in the gene encoding
 A:Reference number: S22066
 A:Accession: S22066
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-514 <DES>
 A:Cross-references: EMBL:X66483; NID:G7289; PIDN:CAA47110.1; PID:G7290
 R:Kamboj, R.K.; Wong, L.M.; Lam, T.Y.; Siu, C.H.
 J. Cell Biol. 107, 1835-1843, 1988
 A:Title: Mapping of a cell-binding domain in the cell adhesion molecule gp80 of Dictyost
 A:Reference number: A31643; MUID:89034443; PMID:3182938
 A:Accession: A31643
 A:Molecule type: mRNA
 A:Residues: 1-514 <RAM>
 A:Cross-references: GB:M36545; NID:G167803; PIDN:AAA33212.1; PID:G167804
 R:Noegel, A.; Gerisch, G.; Stadler, J.; Westphal, M.
 EMBO J. 5, 1473-1476, 1986
 A:Title: Complete sequence and transcript regulation of a cell adhesion protein from agg
 A:Reference number: A26310
 A:Accession: A26310
 A:Molecule type: mRNA
 A:Residues: 1-215, 'G', 217-514 <NOE>
 A:Cross-references: EMBL:X04004; NID:G7246; PIDN:CAA27634.1; PID:G571860
 R:Wong, L.M.; Siu, C.H.
 Proc. Natl. Acad. Sci. U.S.A. 83, 4248-4252, 1986
 A:Title: Cloning of cDNA for the contact site A glycoprotein of Dictyostelium discoideum
 A:Reference number: A23951
 A:Accession: A23951
 A:Molecule type: mRNA
 A:Residues: 20-49 <WON>
 A:Cross-references: GB:M13546; NID:G167699; PIDN:AAA33181.1; PID:G167700
 R:Kamboj, R.K.; Garlepy, J.; Siu, C.H.
 Cell 59, 615-625, 1989
 A:Title: Identification of an octapeptide involved in homophilic interaction of the cell
 A:Reference number: A33463; MUID:90058644; PMID:2582489
 A:Accession: A33463
 A:Molecule type: protein
 A:Residues: 132-139 <KAZ>
 R:Stadler, J.; Bordier, C.; Lottspeich, F.; Henschen, A.; Gerisch, G.
 Hoppe-Seyler's Z. Physiol. Chem. 363, 771-776, 1982
 A:Title: Improved purification and N-terminal amino acid sequence determination of the c
 A:Reference number: S01473; MUID:83005404; PMID:7118072
 A:Accession: S01473
 A:Molecule type: protein
 A:Residues: 20-37, 'X', 39-41, 'X', 43-46 <STA>
 R:Stadler, J.; Keenan, T.W.; Bauer, G.; Gerisch, G.
 EMBO J. 8, 371-377, 1989
 A:Title: The contact site A glycoprotein of Dictyostelium discoideum carries a phospholi
 A:Reference number: A56857; MUID:89251561; PMID:2721485
 A:Contents: annotation; detection of glycosphingolipidinositol anchor

A>Note: no phosphoserine was detected but phosphoinositol and ceramide were
 A>Note: the N-linked carbohydrates are sulfated type 1 oligosaccharides
 C:Comment: This protein is involved in the formation of intercellular contacts upo
 C:Genetics:
 A:Gene: csa
 C:Keywords: blocked carboxyl end; cell adhesion; glycoprotein; lipoprotein; membra
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-514/Product: cell adhesion 80K glycoprotein #status predicted <MAT>
 F:132-139/Region: cell adhesion #status predicted
 F:460-479/Region: 10-residue repeats
 F:493-514/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:128,137,207,294,399/Binding site: carboxylate (Asn) (covalent) #status predicte
 F:492/Modified site: GSI-anchor ethanolamine amidated carboxyl end (Ser) (in matur

Query Match 66.7%; Score 38; DB 2; Length 514;
 Best Local Similarity 70.0%; Pred. No. 25;
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 AGGILHLELL 11
 |||||
 DB 165 AGGLNLELI 174

RESULT 4
 A44100
 cell adhesion molecule gp80 - slime mold (Dictyostelium discoideum)
 C:Species: Dictyostelium discoideum
 C>Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994
 C:Accession: A44100
 R:Desbarats, L.; Lam, T.Y.; Wong, L.M.; Siu, C.H.
 J. Biol. Chem. 267, 19655-19664, 1992
 A:Title: Identification of a unique camp-response element in the gene encoding the
 A:Reference number: A44100; MUID:92406924; PMID:1326559
 A:Contents: AX2
 A:Accession: A44100
 A>Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-514 <DES>
 A>Note: sequence inconsistent with the nucleotide translation
 A>Note: sequence extracted from NCBI backbone (NCBI:P113993)

Query Match 66.7%; Score 38; DB 2; Length 514;
 Best Local Similarity 70.0%; Pred. No. 25;
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 AGGILHLELL 11
 |||||
 DB 165 AGGLNLELI 174

RESULT 5
 S43561
 YCR37C homolog K08E5.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 07-Sep-1994 #sequence_revision 10-Nov-1995 #text_change 24-Nov-1999
 C:Accession: S43561
 R:Kerhaw, J.
 submitted to the EMBL Data Library, March 1994
 A:Reference number: S43561
 A:Accession: S43561
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-526 <KER>
 A:Cross-references: EMBL:Z30974
 C:Genetics:
 A:Introns: 33/3; 122/2; 169/2; 214/2; 253/2; 481/3
 C:Superfamily: sodium/sulfate cotransporter

Query Match 66.7%; Score 38; DB 2; Length 526;
 Best Local Similarity 80.0%; Pred. No. 26;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 GGILHLELLV 12

```

Db      485 GGIISLELLV 494
      |||: |||||
RESULT 6
G88575
protein K08E5.2 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: G88575
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological processes
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: G88575
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-534 <STO>
A:Cross-references: GB:chr_III; PIDN:CAA83225.1; PID:G3878357; GSPDB:GN00021; CESP:K08E5.2
C:Genetics:
A:Gene: K08E5.2
A:Map position: 3
C:Superfamily: sodium/sulfate cotransporter

Query Match      66.7%; Score 38; DB 2; Length 534;
Best Local Similarity 80.0%; Pred. No. 26;
Matches      8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      3 GGIHLLELLV 12
      |||: |||||
Db      493 GGIISLELLV 502

RESULT 7
T31094
surface antigen BspA - Bacteroides forsythus
C:Species: Bacteroides forsythus
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31094
R:Sharma, A.; Sojar, H.T.; Glurich, I.; Honma, K.; Kuramitsu, H.K.; Genco, R.J.
Infect. Immun. 66, 5703-5710, 1998
A:Title: Cloning, expression, and sequencing of a cell surface antigen containing a leucine-rich repeat
A:Reference number: Z20977; MUID:99043895; PMID:9826345
A:Accession: T31094
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1081 <SHA>
A:Cross-references: EMBL:AF054892; NID:G3005672; PID:G3005673; PIDN:AAC82625.1
C:Genetics:
A:Gene: bspA

Query Match      66.7%; Score 38; DB 2; Length 1081;
Best Local Similarity 80.0%; Pred. No. 58;
Matches      8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 AAGGILHLEL 10
      |||: |||||
Db      1023 AAGGALHLLT 1032

RESULT 8
A72502
hypothetical protein APE1991 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C:Accession: A72502
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahashi, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kawauchi, A.; Hara, T.; Kikuchi, Y.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix strain K1
A:Reference number: A72450; MUID:99310339; PMID:10382966

```

Db 46 GGVHLSEJLL 55
|||||

RESULT 11

D70971
hypothetical protein Rv3365c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: D70971
R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: D70971
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-876 <COL>
A:Cross-references: GB:AL009198; GB:AL123456; NID:g3242262; PIDN:CAA15750.1; PID:e120228
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3365c

Query Match 64.9%; Score 37; DB 2; Length 876;
Best Local Similarity 75.0%; Pred. No. 72;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGILHL 8
|||||

Db 520 AAGGVVHL 527
|||||

RESULT 12

T35594
hypothetical protein SC6G4.43c SC6G4.43c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
C:Accession: T35594
R:Saunders, D.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z21593
A:Accession: T35594
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-233 <SAU>
A:Cross-references: EMBL:AL031317; PIDN:CAA20421.1; GSPDB:GN00070; SCOEDB:SC6G4.43c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC6G4.43c
C:Superfamily: hypothetical protein HI0278

Query Match 63.2%; Score 36; DB 2; Length 233;
Best Local Similarity 72.7%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
|||||

Db 211 AAGGALHPEAL 221
|||||

RESULT 13

A44167
aminomethyltransferase (EC 2.1.2.10) precursor - chicken
N:Alternate names: glycine cleavage system protein T
C:Species: Gallus gallus (chicken)
C>Date: 27-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
C:Accession: A44167; B23707
R:Okamura-Ikeda, K.; Fujiwara, K.; Motokawa, Y.
J. Biol. Chem. 267, 18284-18290, 1992
A:Title: Molecular cloning of a cDNA encoding chicken T-protein of the glycine cleavage

translational initiation region on expression.
A:Reference number: A44167; MUID:92406724; PMID:1526969
A:Accession: A44167
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-392 <OKA>
A:Cross-references: GB:D11162; NID:g222867; PIDN:BAA01937.1; PID:g222868
A:Experimental source: liver
A:Note: sequence extracted from NCBI backbone (NCBIN:113296, NCBIPI:113297)
A:Note: parts of this sequence, including the amino end of the mature protein, was
R:Okamura-Ikeda, K.; Fujiwara, K.; Yamamoto, M.; Hiraga, K.; Motokawa, Y.
J. Biol. Chem. 266, 4917-4921, 1991
A:Title: Isolation and sequence determination of cDNA encoding T-protein of the gl
A:Reference number: A23707; MUID:91161577; PMID:2002038
A:Accession: B23707
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 179-392 <OK2>
A:Note: part of this sequence was confirmed by protein sequencing
C:Comment: A mutation of the T-protein in humans leads to nonketotic hypoglycin
C:Superfamily: aminomethyltransferase
C:Keywords: mitochondrion; transferase
F:1-16/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F:17-392/Product: aminomethyltransferase #status predicted <NAT>

Query Match 63.2%; Score 36; DB 2; Length 392;
Best Local Similarity 70.0%; Pred. No. 46;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGGILHLEL 10
|||||

Db 154 AAGGDVHLEV 163
|||||

RESULT 14

A48043
ubiquinol-cytochrome-c reductase (EC 1.10.2.2) core protein I - human
N:Alternate names: cytochrome bc1 core I protein
C:Species: Homo sapiens (man)
C>Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 03-Jun-2002
C:Accession: A48043; I52367
R:Hoffman, G.G.; Lee, S.; Christiano, A.M.; Chung-Honet, L.C.; Cheng, W.; Katchma
J. Biol. Chem. 268, 21113-21119, 1993
A:Title: Complete coding sequence, intron/exon organization, and chromosomal loca
A:Reference number: A48043; MUID:94012661; PMID:8407948
A:Accession: A48043
A:Molecule type: DNA
A:Residues: 1-480 <HOF>
A:Cross-references: GB:L16842; NID:g349472; PIDN:AAA20046.1; PID:g515634
R:Islam, M.M.; Tanaka, M.; Suzuki, H.; Torii, K.; Hattori, M.; Ozawa, T.
Biochem. Mol. Biol. Int. 32, 797-805, 1994
A:Title: A complete cDNA sequence for core I protein subunit of human ubiquinol-c
A:Reference number: I52367; MUID:94348430; PMID:8069229
A:Accession: I52367
A:Status: preliminary; translated from GB/EMBL/DBJ

Query Match 63.2%; Score 36; DB 2; Length 480;
Best Local Similarity 63.6%; Pred. No. 57;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
|||||

Db 233 AAGGVEHQQLL 243
|||||

```
RESULT 15
C98300
hypothetical protein AGR_L_2721 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: C98300
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: C98300
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-505 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89925.1; PID:gl5159878; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_2721
A;Map position: linear chromosome
C;Superfamily: phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent

Query Match      63.2%; Score 36; DB 2; Length 505;
Best Local Similarity 58.3%; Pred. No. 61;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAGGILHLELV 12
      ||||: || -|
Db      107 AAGGVAHLMCVV 118

Search completed: March 20, 2003, 12:35:30
Job time : 5.18919 secs
```

GenCore version 5.1.4.p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:32:11 ; Search time 1.45946 Seconds
(without alignments)
341.028 Million cell updates/sec

Title: SEQID_1

Perfect score: 57

Sequence: 1 AAGGILHLELLV 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	70.2	489	1 MPPB_MOUSE	Q9cxt8 mus musculus
2	40	70.2	489	1 MPPB_RAT	O03346 rattus norv
3	38	66.7	489	1 MPPB_HUMAN	O75439 homo sapien
4	38	66.7	489	1 YN22_CAEEL	O21339 caenorhabdi
5	38	66.7	514	1 CSA_DICDI	P08796 dictyosteli
6	37	64.9	378	1 NOBL_THIFE	P20085 thioobacillu
7	36	63.2	392	1 GCST_CHICK	P28337 gallus gall
8	36	63.2	435	1 FIXC_RHISN	O53208 rhizobium s
9	36	63.2	461	1 NNPI_HUMAN	P56182 homo sapien
10	36	63.2	480	1 UCRI_HUMAN	P31930 homo sapien
11	36	63.2	480	1 UCRI_MOUSE	Q9c213 mus musculus
12	36	63.2	1096	1 PULA_KLEAE	P07811 klebsiella
13	36	63.2	1178	1 YNI7_YEAST	P48231 saccharomyc
14	36	63.2	1805	1 RW1_HUMAN	Q92545 homo sapien
15	36	63.2	1829	1 RW1_MOUSE	O70472 mus musculus
16	35	61.4	145	1 UR1_CYPCA	P01146 cyprinus ca
17	35	61.4	189	1 YTRE_LEPBI	P20464 leptospira
18	35	61.4	480	1 UCRI_BOVIN	P31800 bos taurus
19	35	61.4	691	1 EFG2_SYNY3	P74228 synchocyst
20	35	61.4	694	1 EFG_SYNP6	P18667 synchococc
21	35	61.4	697	1 EFG_SPIPL	P13550 spirulina p
22	35	61.4	698	1 EFG_DEIRA	Q9rxk5 deinococcus
23	35	61.4	1090	1 PULA_KLEPN	P07266 klebsiella
24	34	59.6	355	1 Y816_DROME	Q9vaf0 drosophila
25	34	59.6	359	1 FIX2_RHILE	P07748 rhizobium 1
26	34	59.6	396	1 ALR_BRUME	Q8ydc3 brucella me
27	34	59.6	401	1 NUOD_DEIRA	Q9ru89 deinococcus
28	34	59.6	492	1 GVD2_HALN1	Q9hht2 halobacteri
29	34	59.6	492	1 GVD2_HALN2	P33958 halobacteri
30	34	59.6	545	1 ESTB_DROPS	P25726 drosophila
31	34	59.6	608	1 EX5A_ECOLI	P04993 escherichia
32	34	59.6	688	1 EFG_APPPP	Q9zeu4 apple proli
33	34	59.6	700	1 EFG2_RALSO	Q8xrm7 ralstonia s

34	59.6	702	1	EFG_THICU	O50565 thioobacillu
35	59.6	703	1	EFG1_RALSO	Q8xv10 ralstonia s
36	59.6	866	1	NASA_KLEBN	Q06457 klebsiella
37	59.6	1234	1	YNK5_CAEEL	P34578 caenorhabdi
38	57.9	118	1	ATP2_SYNP6	P08443 synchococc
39	57.9	239	1	TN14_HUMAN	O9qvh9 mus musculu
40	57.9	240	1	TN14_MOUSE	O43557 homo sapien
41	57.9	341	1	EFG_STTRA	P29541 streptomyce
42	57.9	363	1	BASS_ECOLI	P08444 escherichia
43	57.9	417	1	YAT1_SYNP6	P08442 synchococc
44	57.9	423	1	YDIN_ECOLI	P76198 escherichia
45	57.9	434	1	OADB_PASMU	Q9c124 pasteurella

ALIGNMENTS

RESULT 1
MPPB_MOUSE
ID MPPB_MOUSE STANDARD; PRT; 489 AA.
AC Q9CXT8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mitochondrial processing peptidase beta subunit, mitochondrial
DE precursor (EC 3.4.24.64) (Beta-MPP) (P-52).
GN PMPCB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX MEDLINE=21085860; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Aehburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Tovo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kohtauki S.,
RA Hayashizaki Y.,
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- FUNCTION: THE MITOCHONDRIAL PROCESSING PEPTIDASE CLEAVES
PRESEQUENCES FROM MITOCHONDRIAL PROTEIN PRECURSORS (BY
SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Release of N-terminal transit peptides from
precursor proteins imported into the mitochondrion, typically with
Arg in position P2.
CC -1- COFACTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF ALPHA AND BETA SUBUNITS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.ebi.ac.uk/ebis/announcements/](http://www.ebi.ac.uk/ebis/announcements)
or send an email to license@ebi.ac.uk).

```

CC -----
DR EMBL; AK013995; BAB29105.1; -
DR MGD; MG1:1920328; 3110004018Rik.
DR InterPro; IPR001431; Peptidase_M16.
DR Pfam; PF00675; Peptidase_M16; 1.
DR PROSITE; PS00143; INSULINASE; 1.
KW Hydrolase; Metalloprotease; Mitochondrion; Transit peptide.
FT TRANSIT 1 45 MITOCHONDRION (POTENTIAL).
FT CHAIN 46 489 MITOCHONDRIAL PROCESSING PEPTIDASE
FT METAL 101 101 ZINC (BY SIMILARITY).
FT ACT SITE 104 104 ZINC (BY SIMILARITY).
FT METAL 105 105 ZINC (BY SIMILARITY).
FT METAL 105 105 ZINC (BY SIMILARITY).
SQ SEQUENCE 489 AA; 54614 MW; 1B594EE0B6FE34A4 CRC64;

Query Match 70.2%; Score 40; DB 1; Length 489;
Best Local Similarity 72.7%; Pred. No. 4.6;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
Db 243 AAGGVCHNELL 253

RESULT 2
MPPB_RAT STANDARD; PRT; 489 AA.
AC Q03346;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mitochondrial processing peptidase beta subunit, mitochondrial
DE precursor (EC 3.4.24.64) (Beta-MPP) (p-52).
GN PMPCB OR MPPB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=93281757; PubMed=8506385;
RA Paces V., Rosenberg L.E., Fenton W.A., Kalousek F.;
RT "The beta subunit of the mitochondrial processing peptidase from rat
RT liver: cloning and sequencing of a cDNA and comparison with a
RT proposed family of metalloproteases."
RL Proc. Natl. Acad. Sci. U.S.A. 90:5355-5358 (1993).
RN [2]
RP SEQUENCE OF 3-489 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93135783; PubMed=8422255;
RA Kitada S., Niidome T., Nagano T., Ogishima T., Ito A.;
RT "Molecular cloning of the smaller subunit (P52) of rat liver
RT mitochondrial processing peptidase."
RL Biochem. Biophys. Res. Commun. 190:289-293 (1993).
RN [3]
RP REVISTONS.
RA Kitada S.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE MITOCHONDRIAL PROCESSING PROTEASE CLEAVES
CC PRESEQUENCES FROM MITOCHONDRIAL PROTEIN PRECURSORS.
CC -!- CATALYTIC ACTIVITY: Release of N-terminal transit peptides from
CC precursor proteins imported into the mitochondrion, typically with
CC Arg in position P2.
CC -!- COFACTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY.
CC -!- SUBUNIT: HETERODIMER OF ALPHA AND BETA SUBUNITS.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L12965; AAA1633.1; -
DR EMBL; D13907; BAA03007.1; -
DR PIR; PC1229; PC1229.
DR MEROPS; M16.003; -.
DR InterPro; IPR001431; Peptidase_M16.
DR Pfam; PF00675; Peptidase_M16; 1.
DR PROSITE; PS00143; INSULINASE; 1.
KW Hydrolase; Metalloprotease; Mitochondrion; Transit peptide.
FT TRANSIT 1 45 MITOCHONDRION (POTENTIAL).
FT CHAIN 46 489 MITOCHONDRIAL PROCESSING PEPTIDASE
FT METAL 101 101 ZINC (BY SIMILARITY).
FT ACT SITE 104 104 ZINC (BY SIMILARITY).
FT METAL 105 105 ZINC (BY SIMILARITY).
FT CONFLICT 167 167 G -> R (IN REF. 2).
FT CONFLICT 312 312 R -> C (IN REF. 2).
SQ SEQUENCE 489 AA; 54318 MW; 98B56A8C62B9F852 CRC64;

Query Match 70.2%; Score 40; DB 1; Length 489;
Best Local Similarity 72.7%; Pred. No. 4.6;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
Db 243 AAGGVCHNELL 253

RESULT 3
MPPB_HUMAN STANDARD; PRT; 489 AA.
AC O75439;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mitochondrial processing peptidase beta subunit, mitochondrial
DE precursor (EC 3.4.24.64) (Beta-MPP) (p-52).
GN PMPCB OR MPPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=98318631; PubMed=9653160;
RA Mao M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-X., Huang Q.-H.,
RA He K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,
RA Wang Y.-X., Chen S.-J., Chen Z.;
RT "Identification of genes expressed in human CD34(+) hematopoietic
RT stem/progenitor cells by expressed sequence tags and efficient
RT full-length cDNA cloning."
RL Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180 (1998).
CC -!- FUNCTION: THE MITOCHONDRIAL PROCESSING PROTEASE CLEAVES
CC PRESEQUENCES FROM MITOCHONDRIAL PROTEIN PRECURSORS.
CC -!- CATALYTIC ACTIVITY: Release of N-terminal transit peptides from
CC precursor proteins imported into the mitochondrion, typically with
CC Arg in position P2.
CC -!- COFACTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY.
CC -!- SUBUNIT: HETERODIMER OF ALPHA AND BETA SUBUNITS.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```
CC -----
DR EMBL: AF054182; AAC39915.1; -.
DR MEROPS; M16.003; -.
DR Genew; HGNC:9119; PMPCB.
DR MIM; 603131; -.
DR InterPro; IPR001431; Peptidase_M16.
DR Pfam; PF00675; Peptidase_M16; 1.
DR PROSITE; PS00143; INSULINASE; 1.
KW Hydrolase; Metalloprotease; Mitochondrion; Transit peptide.
FT TRANSIT 1 45 MITOCHONDRIAL PROCESSING PEPTIDASE
FT CHAIN 46 489
FT METAL 101 101 ZINC (BY SIMILARITY).
FT ACT SITE 104 104 BY SIMILARITY.
FT METAL 105 105 ZINC (BY SIMILARITY).
SQ SEQUENCE 489 AA; 54167 MW; 440A25A3E4486D64 CRC64;

Query Match 66.7%; Score 38; DB 1; Length 489;
Best Local Similarity 72.7%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGGILHLELL 11
   ||||: |||
DB 243 AAGGVSHDELL 253

RESULT 4
YMW22_CAEEL STANDARD; PRT; 498 AA.
AC Q21339;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein K08E5.2 in chromosome III.
GN K08E5.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Kershaw J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RP REVISIONS.
RA Durbin R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -1- SIMILARITY: BELONGS TO THE NADC/P/PHO87 FAMILY OF TRANSPORTERS.
CC NADC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z30974; CAA83225.2; -.
DR WormPep; K08E5.2; CE28598.
DR InterPro; IPR001898; Na_sulf_syp.
DR Pfam; PF00939; Na_sulf_syp; 1.
DR TIGRFAMs; TIGR00785; Gass; 1.
DR PROSITE; PS01271; NA_SULFATE; 1.
KW Hypothetical protein; Transmembrane; Transport.
FT TRANSIT 24 44 POTENTIAL.
FT TRANSMEM 94 114
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 200 220 POTENTIAL.
FT TRANSMEM 261 281 POTENTIAL.
FT TRANSMEM 284 304 POTENTIAL.
```

```
FT TRANSMEM 332 352 POTENTIAL.
FT TRANSMEM 371 391 POTENTIAL.
FT TRANSMEM 428 448 POTENTIAL.
SQ SEQUENCE 498 AA; 54517 MW; FAEDB2FAA694BCB1 CRC64;

Query Match 66.7%; Score 38; DB 1; Length 498;
Best Local Similarity 80.0%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 GGILHLELLV 12
   |||: ||||
DB 457 GGIIULELLV 466

RESULT 5
CSA_DICDI STANDARD; PRT; 514 AA.
AC P08796; P19408;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Contact site A protein precursor (CSA) (Membrane-associated
DE glycoprotein gp80) (Cell adhesion molecule gp80).
GN CSA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RA Noegel A., Gerlach G., Stadler J., Westphal M.;
RT "Complete sequence and transcript regulation of a cell adhesion
RT protein from aggregating Dictyostelium cells.;"
RL EMBO J. 5:1473-1476(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89105362; PubMed=3063296;
RA Su C.-H., Wong L.M., Lam T.Y., Kamboj R.K., Choi A., Cho A.;
RT "Molecular mechanisms of cell-cell interaction in Dictyostelium
RT discoideum.;"
RL Biochem. Cell Biol. 66:1089-1099(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AX2;
RX MEDLINE=92406924; PubMed=1326559;
RA Desbarats L., Lam T.Y., Wong L.M., Siu C.H.;
RT "Identification of a unique CAMP-response element in the gene
RT encoding the cell adhesion molecule gp80 in Dictyostelium
RT discoideum.;"
RL J. Biol. Chem. 267:19655-19664(1992).
RN [4]
RP SEQUENCE OF 20-49.
RA Wong L.M., Siu C.-H.;
RT "Cloning of cDNA for the contact site A glycoprotein of Dictyostelium
RT discoideum.;"
RL Proc. Natl. Acad. Sci. U.S.A. 83:4248-4252(1986).
RN [5]
RP CELL-BINDING DOMAIN.
RX MEDLINE=99034443; PubMed=3182938;
RA Kamboj R.K., Wong L.M., Lam T.Y., Siu C.H.;
RT "Mapping of a cell-binding domain in the cell adhesion molecule gp80
RT of Dictyostelium discoideum.;"
RL J. Cell Biol. 107:1835-1843(1988).
RN [6]
RP GPI-ANCHOR.
RX MEDLINE=89251561; PubMed=2721485;
RA Stadler J., Keenan T.W., Bauer G., Gerlach G.;
RT "The contact site A glycoprotein of Dictyostelium discoideum carries
RT a phospholipid anchor of a novel type.;"
RL EMBO J. 8:371-377(1989).
CC -1- FUNCTION: THIS CELL-SURFACE GLYCOPROTEIN MEDIATES CELL-CELL
CC BINDING VIA HOMOPHILIC INTERACTION.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC THAT CONTAINS A PHOSPHOCERAMIDE MOIETY. SUCH ANCHOR MEDIATES A
```



```

CC FAST AND LONG PERSISTENCE CELL ADHESION OF THE PROTEIN.
CC -I- DEVELOPMENTAL STAGE: RESTRICTED TO THE AGGREGATION STAGE OF
CC DEVELOPMENT OF D.DISCOIDEUM.
CC -I- PTM: ACYLATED, PHOSPHORYLATED ON SERINE AND N-GLYCOSYLATED WITH
CC TWO TYPES OF OLIGOSACCHARIDE CHAINS.
CC -I- MISCELLANEOUS: THE EXPRESSION OF THIS STRINGENTLY REGULATED
CC PROTEIN DURING CELL DEVELOPMENT IS MEDIATED THROUGH CELL-SURFACE
CC CAMP RECEPTORS.
CC -I- SIMILARITY: THE C-TERMINAL REGION CONTAINS CLUSTERS OF PROLINE
CC REGULARLY ALTERNATING WITH A HYDROXYAMINO ACID AND SHOWS
CC SIMILARITY TO THE HINGE REGION OF IMMUNOGLOBULINS. THIS DOMAIN
CC MIGHT ACT AS A SPACER TO ELEVATE SITES ACTIVE IN CELL CONTACT
CC INTO THE EXTRACELLULAR SPACE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04004; CAA27634.1; -
DR EMBL; M36545; AAA33212.1; -
DR EMBL; X66483; CAA47110.1; -
DR PIR; A23951; A23951.
DR PIR; A26310; A26310.
DR PIR; A31643; A31643.
DR PIR; S22066; S22066.
DR PIR; A44100; A44100.
DR DictyDb; DD02005; csaa.
DR InterPro; IPR002909; IPT_TTG.
DR Pfam; PF01833; TIG; 3.
KW Cell adhesion; Glycoprotein; Membrane; Phosphorylation; Repeat;
KW GPI-anchor; Lipoprotein; Signal.
FT SIGNAL 1
FT CHAIN 20 492
FT PROPEP 493 514 CONTACT SITE A PROTEIN.
FT DOMAIN 20 453 REMOVED IN MATURE FORM (POTENTIAL).
FT DOMAIN 454 485 GLOBULAR (POTENTIAL).
FT DOMAIN 462 479 PRO-RICH (HINGE STRUCTURE).
FT REPEAT 462 469 2 X 8 AA REPEATS, PRO-RICH.
FT REPEAT 472 479 1.
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 492 492 GPI-ANCHOR (POTENTIAL).
FT CONFLICT 216 216 V -> G (IN REF. 1).
SQ SEQUENCE 514 AA; 53696 MW; 96A1C7CF42FEC096 CRC64;

Query Match 66.7%; Score 38; DB 1; Length 514;
Best Local Similarity 70.0%; Pred. No. 12;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 AGGILHLELL 11
DB 165 AGGLNLELI 174
|||:||||:
RESULT 6
MOBL_THIFE
ID MOBL_THIFE STANDARD; PRT; 378 AA.
AC P20085;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 01-FEB-1991 (Rel. 17, Last annotation update)
DE Mobilization protein mobL.
CN MOBL.
OS Thobacillus ferrooxidans.
CG Plasmid pTF1.
OC Bacteria; Proteobacteria; gamma subdivision; Acidithiobacillus.

```

```

OX NCBI_TaxID=920;
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 33020;
RX MEDLINE=91125140; PubMed=2280689;
RA Drole M., Zanga P., Lau P.C.K.;
RT "The mobilization and origin of transfer regions of a Thiobacillus
RT ferrooxidans plasmid: relatedness to plasmids RSF1010 and pSC101.";
RL Mol. Microbiol. 4:1381-1391(1990).
CC -I- FUNCTION: THIS PROTEIN IS ESSENTIAL TO PROMOTE THE SPECIFIC
CC TRANSFER OF THE PLASMID IN THE PRESENCE OF CONJUGATIVE PLASMIDS.
CC -I- SIMILARITY: TO MOBILIZATION PROTEIN A OF THE E.COLI PLASMID INCO
CC RSF1010 AND TO THE 45 kDa RELAXATION PROTEIN OF THE S.TYPHIMURIUM
CC PLASMID PSC101.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52699; CAA36927.1; -
DR PIR; S12190; S12190.
DR InterPro; IPR005053; MobA_Mobl.
DR Pfam; PF03389; MobA_Mobl; 1.
KW Mobility protein; Plasmid; Trans-acting factor; Conjugation.
SQ SEQUENCE 378 AA; 42636 MW; 4AAE91530BC0C194 CRC64;

Query Match 64.9%; Score 37; DB 1; Length 378;
Best Local Similarity 60.0%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 GGILHLELLV 12
DB 46 GGVLHSELIL 55
|||:||||:
RESULT 7
GCST_CHICK
ID GCST_CHICK STANDARD; PRT; 392 AA.
AC P2837;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aminomethyltransferase, mitochondrial precursor (EC 2.1.2.10) (Glycine
DE cleavage system T protein) (GCVT).
GN AMT.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92406724; PubMed=1526969;
RA Okamura-Ikeda K., Fujiwara K., Motokawa Y.;
RT "Molecular cloning of a cDNA encoding chicken T-protein of the
RT glycine cleavage system and expression of the functional protein in
RT Escherichia coli. Effect of mRNA secondary structure in the
RT translational initiation region on expression.";
RL J. Biol. Chem. 267:18284-18290(1992).
RN [2]
RP SEQUENCE OF 179-392 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91161577; PubMed=2002038;
RA Okamura-Ikeda K., Fujiwara K., Yamamoto M., Hiraga K., Motokawa Y.;
RT "Isolation and sequence determination of cDNA encoding T-protein of
RT the glycine cleavage system.";
RL J. Biol. Chem. 266:4917-4921(1991).
CC -I- FUNCTION: The glycine cleavage system catalyzes the degradation of
CC glycine.

```



```

RX MEDLINE=99276573; PubMed=10341208;
RA Savino T.M., Batos R., Jansen E., Hernandez-Verdun D.;
RT "The nucleolar antigen Nop52, the human homologue of the yeast
RT ribosomal RNA processing RRP1, is recruited at late stages of
RT nucleogenesis.";
RL J. Cell Sci. 112:1889-1900(1999).
CC -I- FUNCTION: PLAYS A CRITICAL ROLE IN THE GENERATION OF 28S RRNA.
CC -I- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -I- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN FETAL AND ADULT
CC TISSUES.
CC -I- SIMILARITY: BELONGS TO THE NNP-1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U79775; AAC51625.1; -
DR EMBL; AP001752; BAA95542.1; -
DR EMBL; BC000380; AAH00380.1; -
DR EMBL; AL137757; CAB70909.1; -
KW Nuclear protein.
FT CONFLICT 326 326 K -> R (IN REF. 4).
SQ SEQUENCE 461 AA; 52839 MW; C6F204EF4161AEAE CRC64;

Query Match 63.2%; Score 36; DB 1; Length 461;
Best Local Similarity 72.7%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGGILHLELL 11
Db 43 AAGGFTHDELL 53

RESULT 10
UCR1 HUMAN STANDARD; PRT; 480 AA.
AC P31930;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ubiquinol-cytochrome C reductase complex core protein I, mitochondrial
DE precursor (EC 1.10.2.2).
GN UQCRC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=94012661; PubMed=8407948;
RA Hoffman G.G., Lee S., Christiano A.M., Chung-Honet L.C., Cheng W.,
RA Katchman S., Uitto J., Greenspan D.S.;
RT "Complete coding sequence, intron/exon organization, and chromosomal
RT location of the gene for the core I protein of human ubiquinol-
RT cytochrome c reductase.";
RT J. Biol. Chem. 268:21113-21119(1993).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT ASN-301.
RC TISSUE=Fibroblast;
RX MEDLINE=94348430; PubMed=8069229;
RA Islam M.M., Tanaka M., Suzuki H., Torii K., Hattori N., Ozawa T.;
RT "A complete cDNA sequence for core I protein subunit of human
RT ubiquinol-cytochrome c reductase.";
RL Biochem. Mol. Biol. Int. 32:797-805(1994).
RN [3]
RP ERRATUM.
RX MEDLINE=95038519; PubMed=7951059;
RA Islam M.M., Tanaka M., Suzuki H., Torii K., Hattori N., Ozawa T.;

```

```

RL Biochem. Mol. Biol. Int. 33:410-410(1994).
RN ERRATUM.
RX MEDLINE=95072552; PubMed=7981668;
RA Islam M.M., Tanaka M., Suzuki H., Torii K., Hattori N., Ozawa T.;
RL Biochem. Mol. Biol. Int. 33:815-815(1994).
RN [5]
RP SEQUENCE OF 35-45.
RC TISSUE=Liver;
RX MEDLINE=941147969; PubMed=8313870;
RA Hughes G.J., Frutiger S., Paquet N., Pasquali C., Sanchez J.-C.,
RA Tissot J.-D., Bairoch A., Appel R.D., Hochstrasser D.F.;
RT "Human liver protein map: update 1993.";
RL Electrophoresis 14:1216-1222(1993).
RN [6]
RP VARIANT ASN-301.
RX PubMed=10453733;
RA Valnot I., Kaasis J., Chretien D., de Lonlay P., Parfait B.,
RA Munnich A., Kachaner J., Rustin P., Roetig A.;
RT "A mitochondrial cytochrome b mutation but no mutations of nuclearly
RT encoded subunits in ubiquinol cytochrome c reductase (complex III)
RT deficiency.";
RL Hum. Genet. 104:460-466(1999).
CC -I- FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C
CC REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH
CC IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. THIS PROTEIN MAY
CC MEDATE FORMATION OF THE COMPLEX BETWEEN CYTOCHROMES C AND C1.
CC -I- CATALYTIC ACTIVITY: OH(2) + 2 ferricytochrome c = O + 2
CC ferrocycytochrome c.
CC -I- SUBUNIT: B1 COMPLEX CONTAINS 11 SUBUNITS; 3 RESPIRATORY SUBUNITS
CC (CYTOCHROME B, CYTOCHROME C1, RIESKE PROTEIN), 2 CORE PROTEINS AND
CC 6 LOW-MOLECULAR WEIGHT PROTEINS.
CC -I- SUBCELLULAR LOCATION: Mitochondrial inner membrane.
CC -I- CAUTION: DOES NOT SEEM TO HAVE A PROTEASE ACTIVITY AS IT LACK THE
CC ZINC-BINDING SITE.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC tion between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L16842; AAA20046.1; -
DR EMBL; D26485; BAA05495.1; -
DR HSSP; P31800; 1BE3
DR MEROPS; M16_973; -
DR SWISS-2DPAGE; P31930; HUMAN.
DR PHC1-2DPAGE; P31930; -
DR HSC-2DPAGE; P31930; HUMAN.
DR SIENA-2DPAGE; P31930; -
DR Genew; HGNC:12585; UQCRC1.
DR MW; 191328; -
DR InterPro; IPR001431; Peptidase_M16.
DR Pfam; PF00675; Peptidase_M16; 1.
DR PROSITE; PS00143; INSULINASE; FALSE NEG.
DR Mitochondrion; Inner membrane; Electron transport; Respiratory chain;
KW Oxidoreductase; Transit peptide; Polymorphism.
FT TRANSIT 1 34 MITOCHONDRION
FT CHAIN 35 480 UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX
FT VARIANT 301 301 CORE PROTEIN I.
FT S -> N.
FT /FTID=VAR 013629.
SQ SEQUENCE 480 AA; 52618 MW; AA6B152E0181679 CRC64;

Query Match 63.2%; Score 36; DB 1; Length 480;
Best Local Similarity 63.6%; Pred. No. 27;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGILHLELL 11
|||:|:|

```

```

Db 233 AAGGVEHQQLL 243
RESULT 11
UCR1 MOUSE
ID UCR1_MOUSE STANDARD; PRT; 480 AA.
AC Q9CZ13; Q9CWL6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ubiquinol-cytochrome C reductase complex core protein I, mitochondrial
DE precursor (EC 1.10.2.2).
GN UCRC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schraml L.M., Staebli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Havahehizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C
CC REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH
CC IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. THIS PROTEIN MAY
CC MEDIATE FORMATION OF THE COMPLEX BETWEEN CYTOCHROMES C AND C1 (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: QH(2) + 2 ferricytochrome c = Q + 2
CC ferrocycytochrome c.
CC -1- SUBUNIT: BCI COMPLEX CONTAINS 11 SUBUNITS; 3 RESPIRATORY SUBUNITS
CC (CYTOCHROME B, CYTOCHROME C1, RIESKE PROTEIN), 2 CORE PROTEINS AND
CC 6 LOW-MOLECULAR WEIGHT PROTEINS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane (By
CC similarity).
CC -1- CAUTION: DOES NOT SEEM TO HAVE A PROTEASE ACTIVITY AS IT LACK THE
CC ZINC-BINDING SITE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AK013128; BAB28666.1; -
CC DR EMBL; AK010553; BAB27022.1; -
CC HSP; P31800; 1BE3.
CC MGD; MGI:107876; Ugcrc1.
CC InterPro; IPR001433; Peptidase M16.
CC Pfam; PF006675; Peptidase M16; 1.
CC PROSITE; PS00143; INSULINASE; FALSE NEG.
CC Mitochondrion; Inner membrane; Electron transport; Respiratory chain;
KW Oxidoreductase; Transit peptide.
FT TRANSIT 1 34 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 35 480 UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX
FT CORE PROTEIN I.
FT CONFLICT 223 223 N -> H (IN REF. 1; BAB27022).
FT CONFLICT 318 318 C -> Y (IN REF. 1; BAB27022).
FT SEQUENCE 480 AA; 52768 MW; 9C6D480DC9D5E429 CRC64;
Query Match 63.2%; Score 36; DB 1; Length 480;
Best Local Similarity 63.6%; Pred. No. 27;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAGGILHLELL 11
|||||:|:|
Db 233 AAGGVEHQQLL 243
RESULT 12
PULA_KLEAE STANDARD; PRT; 1096 AA.
ID PULA_KLEAE
AC P07811;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pullulanase precursor (EC 3.2.1.41) (Alpha-dextrin endo-1,6-alpha-
DE glucosidase) (Pullulan 6-glucanohydrolase).
GN PULA.
OS Klebsiella aerogenes.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella
OX NCBI_TaxID=28451;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W70;
RX MEDLINE=87194626; PubMed=3155373;
RA Katsuragi N., Takizawa N., Murooka Y.;
RT "Entire nucleotide sequence of the pullulanase gene of Klebsiella
RT aerogenes W70.";
RL J. Bacteriol. 169:2301-2306(1987).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic
CC linkages in pullulan and in amylopectin and glycogen, and the
CC alpha- and beta-limit dextrins of amylopectin and glycogen.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (probable).
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M16187; AAA25124.1; ALT_SEQ.
CC PIR; A26879; A26879.
CC InterPro; IPR000461; Alpha_amylase.
CC InterPro; IPR004193; Isoamylase_N.
CC InterPro; IPR005323; PUD.
CC Pfam; PF00128; alpha-amylase; 1.
CC Pfam; PF02922; isoamylase_N; 1.
CC Pfam; PF03714; PUD; 1.
CC PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
CC Hydroxylase; Glycosidase; Membrane; Lipoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 1096 PULLULANASE.
FT LIPID 20 20 N-ACYL DIGLYCERIDE.
FT ACT_SITE 694 694 BY SIMILARITY.
FT ACT_SITE 723 723 BY SIMILARITY.
FT ACT_SITE 851 851 BY SIMILARITY.
FT SEQUENCE 1096 AA; 119335 MW; FE7D9167CDACFD79 CRC64;

```

```
Query Match      63.2%; Score 36; DB 1; Length 1096;
Best Local Similarity 54.5%; Pred. No. 63;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGTLHLELL 11
DB 484 SASGVTHIELL 494
      :||:|:|:|:|
      :||:|:|:|:|

RESULT 13
YNI7 YEAST
ID YNI7 YEAST STANDARD; PRT; 1178 AA.
AC P48231;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 132.5 kDa protein in TOP2-MKT1 intergenic region.
GN YNL087W OR N2250.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=96310628; PubMed=8740422;
RA Sotter-Mira A., Saiz J.E., Ballesta J.P.G., Remacha M.;
RT "The sequence of a 17,933 bp segment of Saccharomyces cerevisiae
RT chromosome XIV contains the RHO2, TOP2, MKT1 and END3 genes and five
RT new open reading frames."
RL Yeast 12:485-491(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: CONTAINS 3 C2 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X89016; CAA61423.1; -
DR EMBL; Z71363; CAA95963.1; -
DR HSP; P21707; 1BYN.
DR SGD; S0005031; YNL087W.
DR InterPro; IPR000008; C2.
DR Pfam; PF00168; C2; 3.
DR SMART; SM00239; C2; 4.
DR PROSITE; PS50004; C2 DOMAIN 2; 3.
KW Hypothetical protein; Transmembrane; Repeat.
FT DOMAIN 380 463 C2 DOMAIN 1.
FT DOMAIN 652 733 C2 DOMAIN 2.
FT DOMAIN 985 1070 C2 DOMAIN 3.
FT TRANSMEM 103 123 POTENTIAL.
FT TRANSMEM 272 292 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 766 786 POTENTIAL.
FT TRANSMEM 855 875 POTENTIAL.
FT TRANSMEM 1134 1154 POTENTIAL.
SQ SEQUENCE 1178 AA; 132509 MW; 6597DE191C739FA1 CRC64;

Query Match      63.2%; Score 36; DB 1; Length 1178;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGTLHLE 9
DB 1096 GGVTLHLE 1102
      :||:|:|:|
      :||:|:|:|

RESULT 14
```

```
RW1_HUMAN
ID RW1_HUMAN STANDARD; PRT; 1805 AA.
AC Q92545;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE RMI protein (Fragment).
GN RMI OR KIAA0257.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawabayashi Y.,
RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain."
RL DNA Res. 3:321-329(1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D87446; BAA13387.1; -
FT NON_TER 1
SQ SEQUENCE 1805 AA; 197588 MW; 955589214413AB84 CRC64;

Query Match      63.2%; Score 36; DB 1; Length 1805;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGGTLHLEL 10
DB 169 SSGDLHLEL 178
      :||:|:|:|
      :||:|:|:|

RESULT 15
RW1_MOUSE
ID RW1_MOUSE STANDARD; PRT; 1829 AA.
AC O70472;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RMI protein.
GN RMI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BAUB/c; TISSUE=Brain;
RA Wilkinson R., Fitter S., Tschirke D., Simmons A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY PLAY A ROLE IN THE IMMUNE RESPONSE TO VIRAL
CC INFECTION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

DR EMBL; AF060565; AAC15232.1; -.
DR MGD; MGI:1927110; Rwl.
SQ SEQUENCE 1829 AA; 200508 MW; DOB3E209257AFAB9 CRC64;
Query Match 63.2%; Score 36; DB 1; Length 1829;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAGGILHLEL 10
: : : : :
Db 197 SSGDLHLEL 206

Search completed: March 20, 2003, 12:34:54
Job time : 4.45946 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:33:39 ; Search time 1.94595 Seconds
(without alignments)
181.441 Million cell updates/sec

Title: SEQID_1

Perfect score: 57
Sequence: 1 AAGGILHLELV 12

Scoring table: BLOSUM62.
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	70.2	489	2	US-08-895-521-3
2	40	70.2	489	3	US-09-235-218-3
3	38	66.7	489	2	US-08-895-521-1
4	38	66.7	489	3	US-09-235-218-1
5	36	63.2	480	2	US-08-895-521-4
6	36	63.2	480	3	US-09-235-218-4
7	36	63.2	611	3	US-08-904-871-4
8	36	63.2	1096	4	US-09-346-237-6
9	36	63.2	1142	3	US-08-904-871-12
10	35	61.4	1090	4	US-09-346-237-5
11	33	57.9	239	4	US-08-913-014A-2
12	33	57.9	239	4	US-08-913-014A-3
13	33	57.9	240	4	US-08-913-014A-1
14	33	57.9	240	4	US-09-072-993C-4
15	33	57.9	242	4	US-08-913-014A-25
16	33	57.9	743	4	US-09-077-354B-2
17	32	56.1	170	3	US-08-110-663-2
18	32	56.1	170	4	US-08-432-335-2
19	32	56.1	170	4	US-09-614-022-2
20	32	56.1	237	3	US-09-248-335-48
21	32	56.1	343	4	US-09-413-814-99
22	32	56.1	429	4	US-08-965-762-29
23	32	56.1	429	4	US-09-911-927-29
24	32	56.1	429	4	US-09-911-862-29
25	32	56.1	656	4	US-09-605-785-379
26	32	56.1	656	4	US-09-439-313-379
27	32	56.1	656	4	US-09-352-616A-379

Sequence 380, App
Sequence 380, App
Sequence 380, App
Sequence 378, App
Sequence 378, App
Sequence 378, App
Sequence 378, App
Sequence 27, Appli
Sequence 2, Appli
Sequence 118, App
Sequence 118, App
Sequence 118, App
Sequence 4740, Ap
Sequence 6, Appli
Sequence 9, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 9, Appli
Sequence 6, Appli

28 32 56.1 671 4 US-09-605-785-380
29 32 56.1 671 4 US-09-439-313-380
30 32 56.1 671 4 US-09-352-616A-380
31 32 56.1 1719 4 US-09-605-785-378
32 32 56.1 1719 4 US-09-439-313-378
33 32 56.1 1719 4 US-09-352-616A-378
34 31.5 55.3 92 4 US-09-334-601-27
35 31.5 55.3 418 4 US-09-334-601-2
36 31 54.4 24 1 US-08-484-635-118
37 31 54.4 24 2 US-08-484-631-118
38 31 54.4 24 2 US-08-827-570-118
39 31 54.4 273 4 US-09-134-001C-4740
40 31 54.4 366 1 US-08-481-377-6
41 31 54.4 366 3 US-08-827-336-9
42 31 54.4 366 3 US-09-153-733A-6
43 31 54.4 366 4 US-09-389-705-6
44 31 54.4 366 4 US-09-357-905-9
45 31 54.4 366 5 PCT-US94-00666-6

ALIGNMENTS

RESULT 1
US-08-895-521-3
; Sequence 3, Application US/08895521
; Patent No. 5869311
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi C.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
; TITLE OF INVENTION: SUBUNIT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,521
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0343 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 294589
; US-08-895-521-3

Query Match 70.2% Score 40; DB 2; Length 489;

```

Best Local Similarity 72.7%; Pred. No. 16;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
    ||||: |||
Db 243 AAGGVCHNELL 253

RESULT 2
US-09-235-218-3
; Sequence 3, Application US/09235218
; Patent No. 6001629
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
; TITLE OF INVENTION: SUBUNIT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,218
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/895,521
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0343 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 294589
; US-09-235-218-3

Query Match 70.2%; Score 40; DB 3; Length 489;
Best Local Similarity 72.7%; Pred. No. 16;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
    ||||: |||
Db 243 AAGGVCHNELL 253

RESULT 3
US-08-895-521-1
; Sequence 1, Application US/08895521
; Patent No. 5869311
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi

```

```

; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
; TITLE OF INVENTION: SUBUNIT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,521
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0343 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT14
; CLONE: 1593490
; US-08-895-521-1

Query Match 66.7%; Score 38; DB 2; Length 489;
Best Local Similarity 72.7%; Pred. No. 37;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
    ||||: |||
Db 243 AAGGVSHDELL 253

RESULT 4
US-09-235-218-1
; Sequence 1, Application US/09235218
; Patent No. 6001629
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
; TITLE OF INVENTION: SUBUNIT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```



```
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235,218
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/895,521
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0343 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAINOT14
CLONE: 1593490
US-09-235-218-1

Query Match      66.7%; Score 38; DB 3; Length 489;
Best Local Similarity 72.7%; Pred. No. 37;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY      1 AAGGILHELL 11
      |||||:||||
Db      243 AAGGVSHDELL 253

RESULT 5
US-08-895-521-4
Sequence 4, Application US/08895521
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,521
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0343 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
```

```
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1082896
US-08-895-521-4

Query Match      63.2%; Score 36; DB 2; Length 480;
Best Local Similarity 63.6%; Pred. No. 82;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY      1 AAGGILHELL 11
      |||||:||||
Db      233 AAGGVSHDELL 243

RESULT 6
US-09-235-218-4
Sequence 4, Application US/09235218
Patent No. 6001629
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235,218
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/895,521
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0343 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1082896
US-09-235-218-4

Query Match      63.2%; Score 36; DB 3; Length 480;
Best Local Similarity 63.6%; Pred. No. 82;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

OY 1 AAGGILHELL 11
|||:|:|
Db 233 AAGGVEHOLL 243

RESULT 7
US-08-904-871-4
; Sequence 4, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134USO UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:: Sequence
US-08-904-871-4

Query Match 63.2%; Score 36; DB 3; Length 611;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILH 7
|||:|:|
Db 355 AAGGILH 361

RESULT 8
US-09-346-237-6
; Sequence 6, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; EARLIER FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1096
; TYPE: PRT
; ORGANISM: Klebsiella aerogenes
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(1096)
; OTHER INFORMATION: Pullulanase
US-09-346-237-6

Query Match 63.2%; Score 36; DB 4; Length 1096;
Best Local Similarity 54.5%; Pred. No. 2e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGGILHELL 11
|||:|:|
Db 484 SASGVTHVELL 494

RESULT 9
US-08-904-871-12
; Sequence 12, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134USO UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1142
; TYPE: PRT
; ORGANISM: Alga (Mesocricetum)
US-08-904-871-12

Query Match 63.2%; Score 36; DB 3; Length 1142;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILH 7
|||:|:|
Db 355 AAGGILH 361

RESULT 10
US-09-346-237-5
; Sequence 5, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; EARLIER FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1090
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(1090)
; OTHER INFORMATION: Pullulanase
US-09-346-237-5

Query Match 61.4%; Score 35; DB 4; Length 1090;
Best Local Similarity 54.5%; Pred. No. 3e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGGILHELL 11
|||:|:|
Db 474 SASGVTHVELL 484

RESULT 11
US-08-913-014A-2
; Sequence 2, Application US/08913014A
; Patent No. 6235878
; GENERAL INFORMATION:

APPLICANT: Nishi, Kazunori
APPLICANT: Hikichi, Yukiko
APPLICANT: Shintani, Yasuhide
TITLE OF INVENTION: NOVEL FAS LIGAND-LIKE PROTEIN, ITS
TITLE OF INVENTION: PRODUCTION AND USE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: David G. Conlin, Esq.
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & USHMAN, LLP
STREET: 130 Water Street
CITY: Boston,
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913, 014A
FILING DATE: 04-SEP-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/02480
FILING DATE: July 17, 1997
ATTORNEY/AGENT INFORMATION:
NAME: David G. Conlin
REGISTRATION NUMBER: 27, 026
REFERENCE/DOCKET NUMBER: 342/47694
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 239
TYPE: Amino acid
TOPOLOGY: Linear
MOLECULE TYPE: Protein
US-08-913-014A-2
Query Match 57.9%; Score 33; DB 4; Length 239;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 3 GGIHLLE 9
DB 203 GGVVHLE 209
RESULT 12
US-08-913-014A-3
Sequence 3, Application US/08913014A
Patent No. 6235878
GENERAL INFORMATION:
APPLICANT: Nishi, Kazunori
APPLICANT: Hikichi, Yukiko
APPLICANT: Shintani, Yasuhide
TITLE OF INVENTION: NOVEL FAS LIGAND-LIKE PROTEIN, ITS
TITLE OF INVENTION: PRODUCTION AND USE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: David G. Conlin, Esq.
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & USHMAN, LLP
STREET: 130 Water Street
CITY: Boston,
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913, 014A
FILING DATE: 04-SEP-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/02480
FILING DATE: July 17, 1997
ATTORNEY/AGENT INFORMATION:
NAME: David G. Conlin
REGISTRATION NUMBER: 27, 026
REFERENCE/DOCKET NUMBER: 342/47694
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 239
TYPE: Amino acid
TOPOLOGY: Linear
MOLECULE TYPE: Protein
US-08-913-014A-3
Query Match 57.9%; Score 33; DB 4; Length 239;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 3 GGIHLLE 9
DB 203 GGVVHLE 209
RESULT 13
US-08-913-014A-1
Sequence 1, Application US/08913014A
Patent No. 6235878
GENERAL INFORMATION:
APPLICANT: Nishi, Kazunori
APPLICANT: Hikichi, Yukiko
APPLICANT: Shintani, Yasuhide
TITLE OF INVENTION: NOVEL FAS LIGAND-LIKE PROTEIN, ITS
TITLE OF INVENTION: PRODUCTION AND USE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: David G. Conlin, Esq.
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & USHMAN, LLP
STREET: 130 Water Street
CITY: Boston,
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913, 014A
FILING DATE: 04-SEP-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/02480
FILING DATE: July 17, 1997
ATTORNEY/AGENT INFORMATION:
NAME: David G. Conlin
REGISTRATION NUMBER: 27, 026
REFERENCE/DOCKET NUMBER: 342/47694
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 240
TYPE: Amino acid
TOPOLOGY: Linear
MOLECULE TYPE: Protein
US-08-913-014A-1

Query Match 57.9%; Score 33; DB 4; Length 240;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 GGIHLLE 9
||:||||
DB 204 GGIVHLE 210

RESULT 14
US-09-072-993C-4
Sequence 4, Application US/09072993C
Patent No. 6346388
GENERAL INFORMATION:
APPLICANT: Michael R. Brigham-Burke
APPLICANT: Peter R. Young
TITLE OF INVENTION: A METHOD OF IDENTIFYING AGONIST AND
FILE REFERENCE: GH-50030
CURRENT APPLICATION NUMBER: US/09/072,993C
CURRENT FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/055,513
PRIOR FILING DATE: 1997-08-13
PRIOR APPLICATION NUMBER: 60/056,980
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/057,550
PRIOR FILING DATE: 1997-08-29
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 4
LENGTH: 240
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-072-993C-4

Query Match 57.9%; Score 33; DB 4; Length 240;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 GGIHLLE 9
||:||||
DB 204 GGIVHLE 210

RESULT 15
US-08-913-014A-25
Sequence 25, Application US/08913014A
Patent No. 6235878
GENERAL INFORMATION:
APPLICANT: Miehl, Kazunori
APPLICANT: Hikichi, Yukiko
APPLICANT: Splinter, Yasuaki
TITLE OF INVENTION: NOVEL FAS LIGAND-LIKE PROTEIN, ITS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: David G. Conlin, Esq.
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & USHMAN, LLP
STREET: 130 Water Street
CITY: Boston,
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,014A
FILING DATE: 04-SEP-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/02480
FILING DATE: JULY 17, 1997
ATTORNEY/AGENT INFORMATION:
NAME: David G. Conlin
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 342/47694
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 242
TYPE: Amino acid
TOPOLOGY: Linear
MOLECULE TYPE: Protein
US-08-913-014A-25

Query Match 57.9%; Score 33; DB 4; Length 242;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 GGIHLLE 9
||:||||
DB 206 GGIVHLE 212

Search completed: March 20, 2003, 12:36:01
Job time : 3.94595 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using BW model

Run on: March 20, 2003, 12:34:44 ; Search time 1.94595 Seconds
(without alignments)
329.684 Million cell updates/sec

Title: SEQID_1
Perfect score: 57
Sequence: 1 AAGGTLHELIV 12

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 221153

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCT05_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	100.0	12	10	US-09-833-328-1
2	57	100.0	15	10	US-09-833-328-5
3	57	100.0	15	10	US-09-833-328-11
4	57	100.0	148	10	US-09-833-328-4
5	57	100.0	150	10	US-09-833-328-6
6	57	100.0	203	10	US-09-781-080B-2
7	57	100.0	242	10	US-09-978-979-1
8	57	100.0	1120	10	US-09-781-080B-11
9	57	100.0	1416	10	US-09-836-712-2
10	45	78.9	9	10	US-09-833-328-9
11	36	63.2	125	10	US-09-810-264-34
12	36	63.2	210	10	US-09-272-809-11
13	36	63.2	1096	10	US-09-833-435A-6
14	35	61.4	1090	10	US-09-833-435A-5
15	35	61.4	1091	9	US-09-262-126C-4
16	35	61.4	1091	9	US-10-245-803-4
17	34	59.6	144	9	US-10-101-464A-789
18	34	59.6	316	9	US-09-811-284-143
19	34	59.6	330	9	US-09-738-626-6014

ALIGNMENTS

20	34	59.6	442	9	US-10-117-417-6	Sequence 6, Appli
21	34	59.6	442	9	US-10-117-417-16	Sequence 16, Appl
22	34	59.6	347	9	US-10-101-464A-928	Sequence 928, App
23	34	59.6	1002	9	US-10-101-464A-957	Sequence 957, App
24	33	57.9	81	10	US-09-864-761-35584	Sequence 35584, A
25	33	57.9	132	9	US-09-764-868-906	Sequence 906, App
26	33	57.9	132	9	US-09-764-868-1214	Sequence 1214, Ap
27	33	57.9	159	10	US-09-731-872-402	Sequence 402, App
28	33	57.9	208	10	US-09-027-287-39	Sequence 39, Appl
29	33	57.9	208	10	US-09-252-656B-39	Sequence 39, Appl
30	33	57.9	240	10	US-09-027-287-2	Sequence 2, Appli
31	33	57.9	240	10	US-09-252-656B-2	Sequence 2, Appli
32	33	57.9	240	12	US-10-066-209-4	Sequence 2, Appli
33	33	57.9	551	13	US-09-818-143-21	Sequence 4, Appli
34	33	57.9	602	10	US-09-841-132-495	Sequence 21, Appl
35	33	57.9	602	10	US-09-841-132-495	Sequence 495, App
36	33	57.9	743	9	US-09-836-613-565	Sequence 565, App
37	32	56.1	38	10	US-09-864-761-36574	Sequence 36574, A
38	32	56.1	121	10	US-09-864-761-45340	Sequence 45340, A
39	32	56.1	170	9	US-10-006-856A-162	Sequence 162, App
40	32	56.1	170	10	US-09-099-823-22	Sequence 22, Appl
41	32	56.1	170	10	US-09-951-845-2	Sequence 2, Appli
42	32	56.1	284	9	US-10-028-072-242	Sequence 242, App
43	32	56.1	284	9	US-10-121-049-242	Sequence 242, App
44	32	56.1	284	9	US-10-123-904-242	Sequence 242, App
45	32	56.1	284	9	US-10-140-470-242	Sequence 242, App

RESULT 1
US-09-833-328-1
; Sequence 1, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Schefflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Rando
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Gerritsen, Helena E.
; TITLE OF INVENTION: Compositing a Polypeptide Chain with the Amino Acid Sequence
; FILE REFERENCE: 247,00CIP
; CURRENT APPLICATION NUMBER: US/09/833,328
; PRIOR APPLICATION NUMBER: 2001-04-12
; PRIOR FILING DATE: 09/721,254
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 12
; TYPE: PRT
; ORGANISM: human
US-09-833-328-1
Query Match 100.0%; Score 57; DB 10; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00063;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1 AAGGTLHELIV 12
OY |||||
1 AAGGTLHELIV 12
RESULT 2
US-09-833-328-5
; Sequence 5, Application US/09833328

Patent No. US20020136713A1
GENERAL INFORMATION:
APPLICANT: Laemmle, Bernhard
APPLICANT: Schwarz, Hans-Peter
APPLICANT: Scheiflinger, Friedrich
APPLICANT: Antoine, Gerhard
APPLICANT: Kerschbaumer, RandoIf
APPLICANT: Tagliavacca, Luigina
APPLICANT: Zimmermann, Klaus
APPLICANT: Furlan, Miha
APPLICANT: Turecek, Peter
APPLICANT: Geritsen, Helena E.
TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
FILE REFERENCE: 247,00CIP
CURRENT APPLICATION NUMBER: US/09/833,328
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 09/721,254
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 15
TYPE: PRT
ORGANISM: human
US-09-833-328-5

Query Match 100.0%; Score 57; DB 10; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00079;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILHLELV 12
Db 1 AAGGILHLELV 12

RESULT 3
US-09-833-328-11
Sequence 11, Application US/09833328
Patent No. US20020136713A1
GENERAL INFORMATION:
APPLICANT: Laemmle, Bernhard
APPLICANT: Schwarz, Hans-Peter
APPLICANT: Scheiflinger, Friedrich
APPLICANT: Antoine, Gerhard
APPLICANT: Kerschbaumer, RandoIf
APPLICANT: Tagliavacca, Luigina
APPLICANT: Zimmermann, Klaus
APPLICANT: Furlan, Miha
APPLICANT: Turecek, Peter
APPLICANT: Geritsen, Helena E.
TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
FILE REFERENCE: 247,00CIP
CURRENT APPLICATION NUMBER: US/09/833,328
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 09/721,254
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11
LENGTH: 15
TYPE: PRT
ORGANISM: human
US-09-833-328-11

Query Match 100.0%; Score 57; DB 10; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00079;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILHLELV 12
Db 1 AAGGILHLELV 12

RESULT 4
US-09-833-328-4
Sequence 4, Application US/09833328
Patent No. US20020136713A1
GENERAL INFORMATION:
APPLICANT: Laemmle, Bernhard
APPLICANT: Schwarz, Hans-Peter
APPLICANT: Scheiflinger, Friedrich
APPLICANT: Antoine, Gerhard
APPLICANT: Kerschbaumer, RandoIf
APPLICANT: Tagliavacca, Luigina
APPLICANT: Zimmermann, Klaus
APPLICANT: Furlan, Miha
APPLICANT: Turecek, Peter
APPLICANT: Geritsen, Helena E.
TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
FILE REFERENCE: 247,00CIP
CURRENT APPLICATION NUMBER: US/09/833,328
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 09/721,254
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 148
TYPE: PRT
ORGANISM: human
US-09-833-328-4

Query Match 100.0%; Score 57; DB 10; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.0081;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILHLELV 12
Db 1 AAGGILHLELV 12

RESULT 5
US-09-833-328-6
Sequence 6, Application US/09833328
Patent No. US20020136713A1
GENERAL INFORMATION:
APPLICANT: Laemmle, Bernhard
APPLICANT: Schwarz, Hans-Peter
APPLICANT: Scheiflinger, Friedrich
APPLICANT: Antoine, Gerhard
APPLICANT: Kerschbaumer, RandoIf
APPLICANT: Tagliavacca, Luigina
APPLICANT: Zimmermann, Klaus
APPLICANT: Furlan, Miha
APPLICANT: Turecek, Peter
APPLICANT: Geritsen, Helena E.
TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
FILE REFERENCE: 247,00CIP
CURRENT APPLICATION NUMBER: US/09/833,328
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 09/721,254
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 150
TYPE: PRT
ORGANISM: human
US-09-833-328-6

Query Match 100.0%; Score 57; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.0082;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILHLELV 12
|||||
Db 3 AAGGILHLELV 14

RESULT 6
US-09-781-080B-2
; Sequence 2, Application US/09781080B
; Patent No. US20020142439A1
; GENERAL INFORMATION:
; APPLICANT: HOLLOWAY, JIM
; APPLICANT: SHEPPARD, PAUL
; APPLICANT: YAMAMOTO, GAYLE
; TITLE OF INVENTION: Anti-Angiogenic Intestinal Peptides,
; FILE REFERENCE: 99-82
; CURRENT APPLICATION NUMBER: US/09/781,080B
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-781-080B-2

Query Match 100.0%; Score 57; DB 10; Length 203;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILHLELV 12
|||||
Db 2 AAGGILHLELV 13

RESULT 7
US-09-978-979-1
; Sequence 1, Application US/09978979
; Patent No. US20020151702A1
; GENERAL INFORMATION:
; APPLICANT: Racle, Lisa, A.
; Twine, Natalie, C.
; Agostino, Michael, J.
; Wolfman, Neil
; Morris, Elisabeth
; TITLE OF INVENTION: Aggreganase Molecules
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Home Products Corporation
; STREET: One Campus Drive
; CITY: Parsippany
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07054

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/978,979
FILING DATE: 16-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/60/241,469
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Kapinos, Ellen J.
REGISTRATION NUMBER: 32,345
REFERENCE/DOCKET NUMBER: GI 5435P
TELECOMMUNICATION INFORMATION:

TELEPHONE: (973) 660-5000
TELEFAX: (973) 683-4117
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 242 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-978-979-1

Query Match 100.0%; Score 57; DB 10; Length 242;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILHLELV 12
|||||
Db 42 AAGGILHLELV 53

RESULT 8
US-09-781-080B-11
; Sequence 11, Application US/09781080B
; Patent No. US20020142439A1
; GENERAL INFORMATION:
; APPLICANT: HOLLOWAY, JIM
; APPLICANT: SHEPPARD, PAUL
; APPLICANT: YAMAMOTO, GAYLE
; TITLE OF INVENTION: Anti-Angiogenic Intestinal Peptides,
; FILE REFERENCE: 99-82
; CURRENT APPLICATION NUMBER: US/09/781,080B
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 1120
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1120)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-781-080B-11

Query Match 100.0%; Score 57; DB 10; Length 1120;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILHLELV 12
|||||
Db 105 AAGGILHLELV 116

RESULT 9
US-09-836-712-2
; Sequence 2, Application US/09836712
; Patent No. US20010049106A1
; GENERAL INFORMATION:
; APPLICANT: PFIZER INC.
; TITLE OF INVENTION: ADAMTS POLYPEPTIDES, NUCLEIC ACIDS ENCODING THEM, AND
; FILE REFERENCE: PC10851A
; CURRENT APPLICATION NUMBER: US/09/836,712
; CURRENT FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1416
; TYPE: PRT
; ORGANISM: Human
US-09-836-712-2

Query Match 100.0%; Score 57; DB 10; Length 1416;
Best Local Similarity 100.0%; Pred. No. 0.08;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLV 12
|||:|||||
Db 98 AAGGILHLELLV 109

RESULT 10
US-09-833-328-9
; Sequence 9, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Scheiflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Rndolf
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Gerritsen, Helena E.
; TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
; TITLE OF INVENTION: Comprising a Polypeptide Chain with the Amino Acid Sequence AAGG
; FILE REFERENCE: 247.00CIP
; CURRENT APPLICATION NUMBER: US/09/833,328
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 9
; TYPE: PRT
; ORGANISM: human
US-09-833-328-9

Query Match 78.9%; Score 45; DB 10; Length 9;
Best Local Similarity 100.0%; Pred. No. 2e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLE 9
|||:|||||
Db 1 AAGGILHLE 9

RESULT 11
US-09-810-264-34
; Sequence 34, Application US/09810264
; Patent No. US20020076775A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Virginia C.
; APPLICANT: Ramodu, Omolayo O.
; APPLICANT: Hu, Xu
; APPLICANT: Lu, Guihua
; APPLICANT: Zhang, Lingyu
; TITLE OF INVENTION: WRKY Transcription Factors and Methods
; TITLE OF INVENTION: of Use
; FILE REFERENCE: 1183
; CURRENT APPLICATION NUMBER: US/09/810,264
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/190,467
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Zea mays
US-09-810-264-34

Query Match 63.2%; Score 36; DB 10; Length 125;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGILHLELLV 12
|||:|||||
Db 24 AAGGILHLELLV 35

RESULT 12
US-09-272-809-11
; Sequence 11, Application US/09272809
; Patent No. US20020022239A1
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C.
; TITLE OF INVENTION: Phytofluors as fluorescent labels
; FILE REFERENCE: 2500.118US0
; CURRENT APPLICATION NUMBER: US/09/272,809
; CURRENT FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Mcphylb
US-09-272-809-11

Query Match 63.2%; Score 36; DB 10; Length 210;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILH 7
|||:|||||
Db 153 AAGGILH 159

RESULT 13
US-09-833-435A-6
; Sequence 6, Application US/09833435A
; Patent No. US20020081670A1
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/833,435A
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/346,237
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: 60/094,353
; PRIOR FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1096
; TYPE: PRT
; ORGANISM: Klebsiella aerogenes
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(1096)
; OTHER INFORMATION: Pullulanase
US-09-833-435A-6

Query Match 63.2%; Score 36; DB 10; Length 1096;
Best Local Similarity 54.5%; Pred. No. 2.5e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGILHLELL 11
|||:|||||
Db 484 SASGVTHIELL 494


```

RESULT 14
US-09-833-435A-5
; Sequence 5, Application US/09833435A
; Patent No. US20020081670A1
; GENERAL INFORMATION:
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629 200-US
; CURRENT APPLICATION NUMBER: US/09/833.435A
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/346,237
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: 60/094,353
; PRIOR FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1090
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(1090)
; OTHER INFORMATION: pullulanase
US-09-833-435A-5

```

```

Query Match      61.4%; Score 35; DB 10; Length 1090;
Best Local Similarity 54.5%; Pred. No. 3.6e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
   |||:|:|
Db 474 SASGVTHVELL 484

```

```

RESULT 15
US-09-262-126C-4
; Sequence 4, Application US/09262126C
; Publication No. US20030013180A1
; GENERAL INFORMATION:
; APPLICANT: Miller, Brian S.
; APPLICANT: Shetty, Jayarama K.
; TITLE OF INVENTION: Modified Forms of Pullulanase
; FILE REFERENCE: GC396-2
; CURRENT APPLICATION NUMBER: US/09/262,126C
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1091
; TYPE: PRT
; ORGANISM: Klebsiella pneumonia
US-09-262-126C-4

```

```

Query Match      61.4%; Score 35; DB 9; Length 1091;
Best Local Similarity 54.5%; Pred. No. 3.6e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
   |||:|:|
Db 475 SASGVTHVELL 485

```

Search completed: March 20, 2003, 12:36:34
Job time : 3.94595 secs